

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.,  $*\underset{\sim}{u}$ -). Crucially, evidence from Middle Persian and Parthian shows us that  $*\underset{\sim}{u}$ - is preserved until a relatively late date, meaning that the distribution of reflexes of  $*\underset{\sim}{u}$ - 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  $*\underset{\sim}{u} > b$  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

List of Figures	ii
List of Tables	iii
Acknowledgements	v
<b>1 Introduction</b>	<b>1</b>
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
<b>2 Iranian development and diversification</b>	<b>5</b>
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 $*\acute{z}$ > $[\delta]$ > $d$ . . . . .	12
2.4.2.1.3 PIr $*\acute{s}u$ > $s$ . . . . .	12
2.4.2.1.4 PIr $*\theta r$ > $\zeta$ [s]. . . . .	13
2.4.2.1.5 PIr $*\theta_{\acute{z}}$ > $\acute{s}(i)y$ . . . . .	13
2.4.2.1.6 PIr $*\acute{s}_{\acute{z}}$ > $\theta(i)y$ . . . . .	14
2.4.2.1.7 PIr $*\acute{z}n$ > $(x)\acute{s}n$ . . . . .	14
2.4.2.1.8 PIr $*\acute{z}a$ - rel. pron. $\rightarrow$ OP <i>haya</i> m., <i>taya</i> n., <i>hayā</i> f. . . . .	15
2.4.2.1.9 PIr $*pa\acute{s}\check{c}\check{a}$ - > OP <i>pasā</i> - > MP <i>ps</i> /pas/ > NP <i>pas</i> ‘behind’. . . . .	15
2.4.2.2 Less straightfoward Southwest Iranian-defining innovations . . . . .	15
2.4.2.2.1 PIr $*\acute{s}t$ > $st?$ . . . . .	15
2.4.2.2.2 PIr $*\acute{s}r$ > $\zeta?$ . . . . .	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 <i>*b(r)-, *d-, *g-</i> . . . . .	18
2.5.4.2	PIr <i>*-ft-, *-xt-</i> . . . . .	19
2.5.5	East vs. West Iranian lexis . . . . .	21
2.5.5.1	PIr <i>*kaufa-</i> ‘mountain’ . . . . .	21
2.5.5.2	Real East-West lexical isoglosses? . . . . .	22
2.5.5.3	East Iranian . . . . .	22
2.5.5.3.1	<i>*maiθā-</i> ‘day’ . . . . .	22
2.5.5.3.2	<i>*kapā-</i> ‘fish’ . . . . .	23
2.5.5.4	West Iranian . . . . .	23
2.5.5.4.1	<i>*hižuan-</i> (against other Iranian <i>*hižu(ā)-</i> ) ‘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 <i>*θ</i> . . . . .	25
2.6.2	West Iranian treatment of <i>*θn</i> . . . . .	25
2.6.3	Other West Iranian phonological patterns . . . . .	27
2.6.3.1	Reflexes of PIr <i>*ś</i> , etc. . . . .	27
2.6.3.2	PIr <i>*r</i> + coronal clusters . . . . .	29
2.6.3.3	PIr <i>*u-</i> . . . . .	30
2.6.3.4	West Iranian functional items . . . . .	32
2.7	Conclusion . . . . .	33
<b>3</b>	<b>Computational classification of the Iranian Languages</b> . . . . .	<b>34</b>
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 <i>*θ</i> . . . . .	39
3.2.1.1.3	PIr <i>*śr</i> > <i>sr</i> , <i>ç/s</i> , <i>š</i> , <i>rs</i> . . . . .	39
3.2.1.1.4	PIr <i>*śu</i> > <i>sp</i> . . . . .	40
3.2.1.1.5	Lateralization of dentals . . . . .	40
3.2.1.1.6	PIr <i>*-ft-, *-xt-</i> > <i>-vd-, -yd-</i> . . . . .	40

3.2.1.1.7	Presence of retroflexion . . . . .	40
3.2.1.1.8	Fortition of PIr fricatives . . . . .	41
3.2.1.2	Morphosyntactic Characters . . . . .	41
3.2.1.2.1	Collective <i>*-tā</i> . . . . .	41
3.2.1.2.2	3rd plural in <i>r</i> . . . . .	41
3.2.1.2.3	Causative in <i>*-āyāia-</i> . . . . .	41
3.2.1.2.4	The <i>ezafe</i> clitic . . . . .	41
3.2.1.2.5	The <i>mē-</i> imperfective . . . . .	42
3.2.1.2.6	Preverbal <i>bi-/be-</i> subjunctive/perfective . . . . .	42
3.2.1.2.7	Object marking with <i>rā</i> . . . . .	42
3.2.1.2.8	The 3rd person singular pronominal enclitic . . . . .	42
3.2.1.2.9	Constituent Order . . . . .	42
3.2.2	Lexical Characters . . . . .	42
3.2.2.1	Loans or Inherited Forms? . . . . .	46
3.2.2.1.1	PIr <i>*aiyaka-</i> ‘one’ > <i>yāk</i> . . . . .	46
3.2.2.1.2	PIr <i>*rāšta-</i> ‘right’ > <i>rāst</i> . . . . .	46
3.2.2.1.3	PIr <i>*hakaram</i> > <i>agar</i> ‘if’ . . . . .	47
3.3	Results . . . . .	47
3.3.1	Non-lexical characters . . . . .	47
3.3.2	Lexical characters . . . . .	48
3.3.2.1	Languages of all chronological periods . . . . .	49
3.3.2.1.1	Loans coded . . . . .	49
3.3.2.1.2	Loans uncoded . . . . .	49
3.3.2.2	Contemporary languages . . . . .	50
3.3.2.2.1	Loans coded . . . . .	50
3.3.2.2.2	Loans uncoded . . . . .	51
3.3.2.2.3	Distance-based modeling of lexical character data . . . . .	52
3.3.2.3	Summary . . . . .	53
3.4	Discussion and Conclusion . . . . .	54
<b>4</b>	<b>Phonological reduction and grammatical convergence in West Iranian</b>	<b>56</b>
4.1	Introduction . . . . .	56
4.2	West Iranian . . . . .	58
4.3	Irregular Sound Change . . . . .	60
4.4	Historical Phonology of Persian Functional Morphemes . . . . .	60
4.5	Modeling Preliminaries . . . . .	64
4.5.1	Levenshtein Distance . . . . .	64
4.5.2	Mixed-Effects Models . . . . .	65
4.6	Method and Results . . . . .	67
4.7	Phonological Change in West Iranian Lexical Items . . . . .	69
4.8	Discussion and Conclusion . . . . .	71

<b>5</b>	<b>Geography and contact-induced variation in West Iranian</b>	<b>73</b>
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
<b>6</b>	<b>Conclusion</b>	<b>87</b>
	<b>Appendix A: Chapter 3 Supplementary Materials</b>	<b>89</b>
	<b>Appendix B: Chapter 4 Supplementary Materials</b>	<b>121</b>
	<b>Appendix C: Chapter 5 Supplementary Materials</b>	<b>132</b>
	<b>Bibliography</b>	<b>135</b>



# List of Figures

2.1	Map of Iranian Languages . . . . .	6
3.1	NeighborNet with Binary Typological/Recurrent Characters . . . . .	47
3.2	Typological data; consensus tree . . . . .	49
3.3	Lexical data; all languages; loans coded . . . . .	50
3.4	Lexical data; all languages; loans uncoded . . . . .	51
3.5	Lexical data; contemporary languages; loans coded . . . . .	52
3.6	Lexical data; contemporary languages; loans uncoded . . . . .	53
3.7	NeighborNet with Lexical Characters, loans coded . . . . .	54
4.1	Approximate geographic locations of West Iranian languages treated in this chapter . . . . .	59
4.2	Multidimensional scaling of inter-language distances . . . . .	70
5.1	Schema of West Iranian dialects in MacKenzie 1961:71, according to non-varying phonological and morphological features . . . . .	76
5.2	Correlation coefficient $r$ for the relationship between great-circle distance and palatal distance, lateral distance, and labiovelar distance . . . . .	81
5.3	Density curves for each distance type . . . . .	82
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 <i>only</i> NW Iranian languages are light print) . . . . .	83
5.5	Correlation coefficient $r$ for the relationship between geographic distance and overall linguistic distance, palatal distance, lateral distance, and velar distance (excluding reflexes of $*_{\text{r}}ur-$ ) . . . . .	85

# 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

$*s > *h$ , 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  $\bar{i}$ -genitive, which has been linked to derivation of the  $vrk\bar{i}$ -type — ignoring also the fact that  $*-osi\bar{o}$ -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 ming: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,  $*ia\bar{z}ata-$  ‘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.

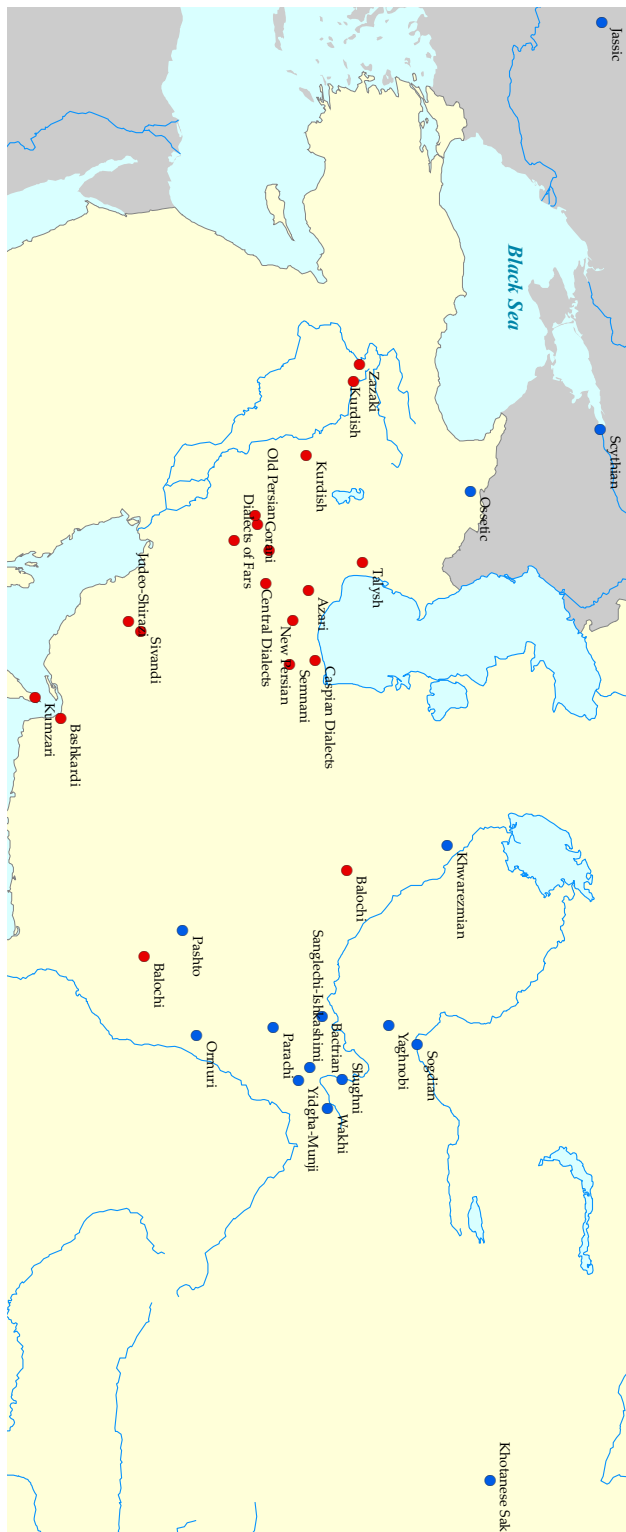
I also sketch out some desiderata for future study of West Iranian similarities, which serve as the basis for this dissertation's remaining chapters.

### 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, Pāzand<sup>1</sup>)
  - iii. New: New (Modern) Persian (NP), Lurī (Lur), Kumzārī (Kmz, Thomas 1930), Baškardī (Bshk, Skjærvø 1988; Voskanian and Boyajian-Sureniantz 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), Zāzākī (Zaz), Gorānī (Gor), Awromānī (Awr, Benedictsens and Christensen 1921) Tālešī (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 Mury 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, Németh 1959), “Parnian” (substrate found in Parthian, Sims-Williams 1989b:171, and possibly Armenian, Olsen 2005)
  - (c) New: Pašto (Psht), Ormuṛī (Orm, Morgenstierne 1929), Parāčī (Par, *ibid.*), Ossetic (Oss, Iron dialect, unless noted as [D]igor, Thordarson 2009), Yaḡnōbī (Ygh, Andreev and Peščereva 1957); so-called “Pamir” languages include Šuḡnī (Shg, Zarubin 1960), Sanglīčī-Iškāšmī (SI), Yidḡā-Munḡ(ān)ī (YM), Yazḡlāmī (Yzg)
- (a) Sakan subgroup
  - i. Middle: Tumšūqese (Tshq) and Khotanese (Khot) Saka
  - ii. New: Waxī (Wkh)

---

<sup>1</sup>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  $*(h_1)e-$  to all aorist verbs (also affects Greek and Armenian)
2. LABIO- AND PLAIN VELAR MERGER: PIE  $*k, *k^w > \text{PIIr } *k; *g, *g^w(^h) > *g(^h)$
3. THE LAW OF THE PALATALS: Early PIIr  $*k, *g(^h) > \text{PIIr } *č, *j(^h)$  before (PIE)  $*e, *i$
4. INDO-IRANIAN VOWEL MERGER: PIE  $*e, *a > \text{PIIr } *a; \text{PIE } *o > \text{PIIr } *\bar{a}; \text{PIE } *ē, *ō, *\bar{a} > \text{PIIr } *\bar{a}$
5. BRUGMANN'S LAW: PIE  $*o > \text{PIIr } *\bar{a}$  in an open syllable,  $*a$  in a closed syllable
6. BARTHOLOMAE'S LAW:  $\left\{ \begin{smallmatrix} s \\ t \end{smallmatrix} \right\} \rightarrow \left\{ \begin{smallmatrix} z^h \\ d^h \end{smallmatrix} \right\} / D^h \_ (D^h \text{ represents a segment with breathy voice})$   
(the process  $D \rightarrow T / \_ \left\{ \begin{smallmatrix} s \\ t \end{smallmatrix} \right\}$  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  $D(^h)$ )  $> \text{PIIr } *š, *ž$  after  $*r, *r', *u, *k^{(w)}, *k', *i$
8. SATEM TREATMENT OF THE PIE PALATAL VELARS: PIE  $*k', *g(^h) > \text{PIIr } *č, *j(^h)$
9. LARYNGEAL MERGER: PIE  $*h_1, *h_2, *h_3 > \text{PIIr } *H; \text{PIIr } *CH > *C^h$  (exceptions due to Inner-IE laryngeal developments)
10. LARYNGEAL VOCALIZATION:  $*H > *\underset{\cdot}{H} / (-)C\_C^-, -C\_ (C), -C\_sR-$  (Lipp 2009:II 485-7)
11. PIE  $*n > *an / \_ V, *a / \_ \left\{ \begin{smallmatrix} C \\ \# \end{smallmatrix} \right\}; \text{PIE } *m > *am / \_ R, *a / \_ \left\{ \begin{smallmatrix} O \\ \#? \end{smallmatrix} \right\}^2$

---

<sup>2</sup>Kobayashi (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  $*l, *r > \text{PIIr } r$  (the status of this merger is questionable; we see what appears to be PIE  $*l$  in peripheral IIr dialects); PIE  $*l̥, *r̥ > \text{PIIr } *r̥$
13. PIE  $*oh_xom \rightarrow \text{PIIr } *\bar{V}n\bar{a}m$  in the genitive plural ending for thematic nouns
14. PIE  $*h_1eǵoh_2-$  ‘I’  $\rightarrow *h_1eǵh_2-$  *om*, and other pronominal changes (see Meillet 1922:26–7)
15. PIIr  $*ć > *š / \_ t, t \_$ , possibly  $p \_$
16. PIIr  $*/s/ \rightarrow [š] / \_ č$

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

1. PIIr  $*p^{(h)} *t^{(h)} *k^{(h)} > \text{PIr } *f *θ *x$
2. PIIr  $*p *t *k > \text{PIr } *f *θ *x$  before consonants
3. PIIr  $*s > \text{PIr } *h$ , 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. PIIr  $*b^h *d^h *g^h > \text{PIr } *b *d *g$
5. PIIr  $*t^st^{(h)} > \text{PIr } *st, *d^zd^{(h)} > \text{PIr } *zd$
6. PIIr  $*H > \emptyset$  in non-initial syllables (Lipp 2009:II 485)
7. PIIr  $*ps > \text{PIr } *fš, *bz > \text{PIr } *βž$
8. PIIr  $*ts, *ć > \text{PIr } *ś, \text{PIIr } *j > *ž$  (see below)
9. PIIr  $*ćš > \text{PIr } *š$
10. PIIr  $*(ć)sć > *ś$
11. PIIr  $*-śn- > *-šn-$  (e.g., YAv *ašnō* : OInd *aśnas* gen.sg. ‘stone’),  $*śn- > *sn-$  (e.g., YAv *snaθ-* : OInd *śnath-* ‘strike’)

Authors represent the Proto-Iranian palatals ( $< \text{PIIr } *ć, *j$ ) according to different conventions. Some use affricates (e.g.,  $*ts, *dz; *ć, *j$ ; an advantage of this convention is that PIIr  $*ts$  behaves identically to PIIr  $*ć$  in Iranian, cf. the separate developments of the Avestan and Persian word for ‘fish’: PIIr  $*mats̥ia-$   $> \text{PIr } *māts̥ia-$   $> \text{YAv } mas̥iia-$ , OP  $*māθiya(ka)- > \text{MP } m̥hy(g) /ma:hi:g/$  : Ved. *matsyā-*), believing Old Persian  $θ$  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  $*ću, *ju > *š, *ž$  shows that these sounds were still alveopalatal in PIr; however, the low F1–3 of [w] could coarticulatorily lower the spectral energy of preceding

[s], yielding a secondary [ç] or [ʃ]. Some authors even believe that a PIr affricate \*č [tç] was preserved in some Old Iranian dialects in some contexts. Tocharian A *añčwāši* ‘in steel’, B *eñcuwo*, *iñcuwo* ‘steel’ are most definitely Iranian loans, perhaps from “Old Sakan” \**anč̣uān-* (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* /antʃa:j/ ‘rest’ < \**han-č̣iāia-* and related forms, which have not undergone the change \*č̣ > š(y), vs. CSog *ptš’dy* < \**pati-č̣iātaka-*).<sup>3</sup> 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 \*š and \*ž to minimize confusion with other sibilants.

Voiceless consonants spirantize before other consonants; however, Avestan shows *pt* as a reflex of PIIr \**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.

- PIr \*ṣ̌u > š: \**aṣ̌uana-* > Khot *hāśśana-* /hi:ʃ(:)ana/ ‘iron’, Wkh *yišn* ‘iron’; \**uīṣ̌ua-* > Tum *biša-*, Khot *biśśa-* /βiʃ(:)a/ ‘all’; \**aṣ̌ua-* > Khot *aśśa-* /aʃ(:)a/, Wkh *yaš* ‘horse’. The Ossetic word *sist*, D *sistæ* ‘flea’ (probably < \*ṣ̌iša-) is ascribed by Thordarson

<sup>3</sup>Sims-Williams (1989d:261) attributes Sogdian *c* in *’nc’y* to a dissimilation \*č̣i...i > č...i. This account may also explain Brahmi Sogdian *icā-t* ‘comfortable’ < \**ūžč̣(i)āta-* (Sims-Williams 1996:308), if formed on the basis of a dissimilated stem \**ūžč̣āia-* < \**ūžč̣iāia-*.

(2009:13) to contact with a “cognate dialect,” probably of a Sakan affiliation (cf. Wkh *šiš* ‘*id.*’ < \**šuiš-*).

- PIr \**ž*u > *ž* (?): \**hižua-ka-* ‘tongue’ > Khot *bišāa-* /βiʒaː/, Wkh *zik* (with secondary depalatalization?); \**žuar-* > Late Khot *šī’r-* /ʒir/ ‘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 PIIr \**č*, \**ǰ*, given the ability of [w] to palatalize an adjacent [s] or [z] (discussed above). There is no Tumshuqese evidence for the change \**ž*u > *ž*, and the idea that Wakhi depalatalizes an earlier \**ž* 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., Fārsī, Tājīk, Hazāragī, Aimaq, Tat-Persian) as well as Kumzārī, Baškārdī, Lārestānī, Bandārī, dialects of Fars, and Lurī (including the dialects Būyer-Ahmadī, Mamasanī, Kohgilūya, Baxtiyārī, and perhaps Lakī). 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).<sup>4</sup>

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 *σπάλα* ‘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.

---

<sup>4</sup>For instance, in many situations, we can see the breakdown of the Old Persian case system as it leads into Middle Persian, e.g., *būmim* ‘earth’ ACC (Darius, Naqš-i Rostam a) → *būmām* ‘*id.*’ (Artaxerxes, Persepolis a) > MP *bwm* /bu:m/. However, elsewhere, 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 \*ś > θ.** At the same time, a number of words found in Persian show *s* for PIr \*ś, usually ascribed to “Median” or other NW Iranian admixture. Word-medially, OP θ > MP *h* > NP *h*. (Non-Persian Southwest Iranian dialects show the change \*-θ- > -*h* as well.) Word-initially, the picture is unclear. OP θ- corresponds to MP, NP *s*- across the board, with at least one exception, \*śata-pād(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 Šīrāzī *tanzīdan* ‘weigh’, a cognate of NP *sanjīdan* ‘id.’, is such a dialectal form, connecting it to OP *aθangainā/i-* ‘stony’). Either (i) the MP and NP forms are NW Iranian borrowings, or (ii) there was a change of the type PIr \*ś- > OP θ- > MP *s*-; the latter scenario seems to be the case, given forms like PIr \*θaxta- > NP *saxt* ‘hard’ (cf. Khw *θyd*), and if PIr \*taigra-(či-) > Scythian \*θaigra-(či-) → OP month name *θaigra(cī-)* (as per Lubotsky 2002:199) > NP *sīr* ‘garlic’. This correspondence was noted by Salemann (1901:263-264). If PIr \*aśanga- > OP *aθanga-* regularly yields MMP *sng /sang/* > NP *sang* (and it is not a NWIr loan), then a change of OP θ- > MP *s*- may postdate the aphaeresis of unstressed vowels.

**2.4.2.1.2 PIr \*ž > [ǰ] > d.** The same caveats as above apply. The regular reflex appears in the following words:

- \*ažam ‘I’ > OP *adam*
- \*iž- ‘worship, sacrifice’ > OP *yad-*
- \*žrd- ‘heart’ > MP *dyl /dil/*, NP *dil*
- \*žrāduni- (cf. Ved *hrāduni-*) ‘hail’ > SBshk *dārāyen*
- \*žraiah- ‘sea’ > OP *drayah-*, MP *dry’(b) /draja:b/*, NP *daryā* ‘river’

However, PIr \*ž > *z* in the following:

- \*iž- ‘worship, sacrifice’ > MMP *yz- /yaz/*
- \*užra- > NP *gurz* ‘mace’

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

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

- \*aśua-bāra- ‘rider’ > OP *asabāra-* > MP *’sw’r* > NP *savār*
- \*śuaka- ‘dog’ > MP *sg /sag/*, NP *sag*, Kms *sōy*, Frs (Pāpuni) *sag*, Lur (Feyli) *sag*
- \*śuiš- ‘louse’ > (Pre-Kms *sōš*) > Kms *šōš*
- \*gau-śuanta- ‘holy cow’ > Kms *yosen* ‘sheep’



\*ś<sub>u</sub>antā 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*:

\*aś<sub>u</sub>a- ‘horse’ > OP *aspa-*, NP *asb*

\*ś<sub>u</sub>iš- ‘louse’ > NP *sepeš*

\*ś<sub>u</sub>anta- ‘holy’ > NP *espan* ~ *sipand* ~ *esfand* ‘wild rue’

\*ś<sub>u</sub>antā aramati- ‘holy earth’ > MP *spndrmt* /spandarmad/ (cf. Arm *Sandaramet*)

\*ś<sub>u</sub>aīta- ‘white’ > NP *sipēd* ~ *sifēd*

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

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

\*θraīa- ‘three’ > MMP *sh* /se(ɾ)/, NP *sih*

\*ā-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 (*M<sup>i</sup>θra* ~ *M<sup>i</sup>tra*), 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:

\*ś<sub>u</sub>iθra- (cf. Ved *śvitrá-* ‘white’) > Phl *spy* /spihr/ > NP *sipihr* ‘sky, firmament’ (if not from Greek σφαῖρα ‘sphere, globe’)

\*cīθra- > Phl *cyhl* /tʃihr/ > NP *čih* ‘form, face’ (cf. Elam *Ti-iš-šá-an-tam-ma* = \**Tiçantama-* < OP PN *Ciça<sup>n</sup>taxma-* < \**čihrantaxma-*)

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

---

<sup>5</sup>There 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) \**θyāva-* ‘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 \**śi-*.

**2.4.2.1.5 PIr**  $*\theta_{\bar{i}} > \check{s}(i)y$ . This is a fairly solid Persian innovation, as seen in the following examples; the rest of Iranian shows no palatalization of  $*\theta$  before a glide (though similar-looking palatalizations can be seen in far-flung East Iranian languages, e.g.,  $*hu_{\bar{a}}i_{\bar{a}}-pa\theta_{\bar{i}}i_{\bar{a}}ka-$   $>$  *xbsk* ‘own’, Humbach 1989:195,  $*a\theta_{\bar{i}}i_{\bar{a}}-(ka)-$  ‘duck’  $>$  Yidyā *yeṣko*, Rastorgueva and Édel’man 2003:I).

$*ha\theta_{\bar{i}}iam >$  OP *hašiyam* ‘truth’ acc. sg.

$*hu_{\bar{a}}i_{\bar{a}}-pa\theta_{\bar{i}}i_{\bar{a}}- >$  Phl *NPŠE*, *hwysš*, MMP *xwy(b)š* /xwe:ʃ/ ‘self’

$*pa\theta_{\bar{i}}i_{\bar{a}}-$  (Skjærvø 1989:364) ‘before’  $>$  MP *pyš* /pe:ʃ/  $>$  NP *peš*

It is not clear if non-Persian Southwest Iranian languages undergo palatalization; these languages tend to show  $\emptyset$  ( $<$   $*h?$ ), e.g.,  $*pa\theta_{\bar{i}}i_{\bar{a}}-$   $>$  NBshk *pī* ‘before’, Kmz *pī* ‘from’ (?) (Skjærvø 1989:364); this cannot be secondary  $<$   $*š$ , given forms like Kmz *šöš* ‘louse’  $<$   $*š_{\bar{u}}i_{\bar{a}}š-$ , which preserve  $*š$  as such.

**2.4.2.1.6 PIr**  $*š_{\bar{i}} > \theta(i)y$ . The fate of PIr  $*š_{\bar{i}}$  is somewhat up in the air in Southwest Iranian. Old Persian inscriptions show  $-\theta i y-$  (e.g., *viθiyā* ‘house’ loc. sg., *fraθiya-* ‘punish’ passive stem; Kent 1951:34 calls *pašiyā-* ‘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-e-na*  $\sim$  *Ši-ia-a-na*  $\sim$  *Ši-ia-e-na* = OP (?)  $*š_{\bar{y}}aina-$  ‘eagle’, Elamite *Ši-ia-ma* = OP  $*š_{\bar{y}}āva-$  ‘black’, alongside Elamite *Ti-ia-ma* = OP  $*\theta_{\bar{y}}āva-$  ‘black’, an “archaic equivalent of  $*š_{\bar{y}}āva-$ ” (Tavernier 2007:316, 319, 330).

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

- PIr  $*š_{\bar{i}} > s$ :  $*š_{\bar{i}}(i)aina-mr̄ga-$  (cf. YAv *saēna-* ‘eagle’, with dissimilation of the first glide?)  $>$  Phl *synmwlw* /se:n murw/ ‘a fabulous bird’, NP *sīmurγ* (MacKenzie 1971a:74);  $*kaš_{\bar{i}}iapa-$   $>$  Bandarī *kāsapošt* ‘turtle’ (+ *pošt* ‘back’, with perhaps haplology) (Rastorgueva and Édel’man 2003:IV 338);
- PIr  $*š_{\bar{i}} > s(i)y$ :-  $*š_{\bar{i}}āua-$   $>$  MP *sy’h*  $>$  NP *syāh* ‘black’
- PIr  $*š_{\bar{i}} > *\theta(i)_{\bar{i}} > h$ :  $*māš(i)_{\bar{i}}i_{\bar{a}}ka-$  ‘fish’ (Hoffmann (1976b:637, fn. 25) ascribes the long vowel to Vrddhi)  $>$  Phl *m’hyg* /ma:hi:g/  $>$  NP *māhī*, Band *mūyi*, Kmz *mī*, Lar *ma’i*; PIIr  $*tus-sć(i)_{\bar{i}}i_{\bar{a}}ka >$  MMP *twhyg*/*twhyg* /*tuhi:g*/ ‘barren’

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

**2.4.2.1.7 PIr**  $*žn > (x)šn$ . From what evidence we have, it looks as though PIr  $*žn$  became  $-šn-$  word-medially (e.g.,  $*i_{\bar{a}}žna-$   $>$  Phl *yšn* /dʒaʃn/  $>$  NP *jašn* ‘festival’,  $*gažna-$   $>$  NP *gašn* ‘abundance’) and  $xšn-$  word-initially (e.g.,  $*žnā-sa-$   $>$  OP *xšnāsa-* ‘know’ inch.  $>$  MMP  $*šn’s-$  /iʃna:s/, Phl  $*šn’s-$  /ʃna:s/  $>$  NP *šinās-* ‘recognize’). Some forms, likely loans from Northwest or East Iranian, contain *zn*, e.g., NP *gavazn* ‘deer’ (cf. MSog  $\gamma wzn-$ , Khot *ggūyzna-*  $<$  PIr  $*gau-až-na-$ ).

The word-medial outcome of *\*-žn-* distinguishes SWIr from NWIr, but from what evidence we have, initial *\*žn-* > *(x)šn-* in NWIr, e.g., *\*žnā-sa-* > *\*(x)šnā-sa-* > Pth *šn's-* /iʃna:s/ ‘recognize, get to know’ inch., Awr *ažnās-* (with voicing), Gil *šenās-*, Xunsārī *išnās-*, Qohrūdī *ešnās-* (Cheung 2007:466).

Avestan shows word and morpheme-initial *xšn-*, e.g., *\*fra-žnin-* > YAv *frašnin-* ‘prescient (?)’, as well as word-initial *žn-* (e.g., *\*žnu-* > YAv *žnu-* ‘knee’, *\*žnā-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-* < *\*iž-na-*) due to analogy. Schwartz (2010) proposes that *-sn-* is the elsewhere condition of the regular outcome, with *-šn-* found in rhotic environments.

Initial *\*žn-* yields Khw *ñ-*, e.g., *\*žnā-sa-* > *m/ñ's* /n:a:s/ (?) ‘become acquainted, accustomed’ inch.; cf. *\*xšnauθra-* ‘that which reciprocates’ (?) > *xnwry* ‘thank’ (MacKenzie 1971b:532; Cheung 2007:467). These forms serve as evidence that *\*žn-* and *\*xšn-* do not merge in Khwarezmian (cf. Pth *šnwyr* → Armenian *šnorh* ‘thanks’).

Sogdian shows initial (or at least stem-initial) *\*žn-* > *(x)šn-* (though *\*žnāka-* > BSog *žn'kh* ‘knowledge, jñāna’ goes against this generalization), e.g., *\*fra-žnā-sa-* > MSog *fšn's-* ‘recognize’ (Cheung 2007:467); Sogdian has *-zn-* for medial *\*-žn-*, e.g., *\*gažna-* > MSog *γzn-/yazn/* ‘treasure’ — essentially the same distribution of NWIr.

Khotanese Saka has *-zn-* for medial *\*-žn-*, e.g., *vaysña* /vazɲa/ ‘now’ < *\*aya-žniā-* (Bailey 1979:376). The behavior of initial *\*žn* is unclear.

For New East Iranian, Ormuri shows *n* < *\*-žn-*, e.g., *injān* ‘day before yesterday’ < *\*anija-ažna-*. Some Pamir languages show *-zd-*, e.g., Sgl *āluzd* ‘id.’ < *\*ā-uša-ažna-* (Rastorgueva and Édel'man 2003:I 295).

**2.4.2.1.8 PIr *\*iā-* 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*s are highly reduced, usually to *=ī* or *=e*).

**2.4.2.1.9 PIr *\*paščā-* > OP *pasā-* > MP *ps* /pas/ > NP *pas* ‘behind’.** Tedesco (1945:128, fn. 4) explains OP *kaščiy*, *čiščiy* ‘something, anything’ as NW Iranian loans (analogical 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 *\*žd* and *\*zd*, e.g., *mužd* ~ *muzd* ‘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 *<sup>h</sup>u-frašta-* ‘well punished’ against *<sup>h</sup>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 *\*žauštā-* >

OP *dauštā-* ‘friend’ (> MP *dwst* /do:st/ > NP *dost*), which shows the Southwest Iranian reflex of PIr *\*ž* 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* /angust/ ‘finger’ (< *\*angušta-*, alongside NP *angušt*, as well as Phl *mwst* ~ *mwšt* ‘fist’, NP *mušt* < *\*mušti-*), Phl *b’lyst* /ba:list/, MMP *b’ryst* /barist/ ‘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-śraǰi-* ‘restore’ > OP *nīyaçāray-* caus. (× *dāraya-* ‘hold’ caus., Kent 1951:188), MP *ns’y* /nisa:j/ ‘conveying, dispatch’. Middle and New Persian often show *sr*:

- *\*śrāuǰia-* ‘hear’ (caus.) > *\*śrāuǰia-* > MP *sr’y* /sra:j/ ‘sing’, NP *surū(dan)/sarāy*
- *\*śrauni-* > NP *surūn* ‘buttocks’
- *\*h<sub>u</sub>ǰāśrū-* > NP *xusrū* ‘mother-in-law’
- *\*ǰśra/u-* ‘tear’ > MMP *’rs* /ars/ > NP *ars* (also MMP *’sr*)

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

- *\*h<sub>u</sub>ǰāśrū-* > NP *xušū* ‘mother-in-law’
- *\*ǰśruka-* > 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 *\*śu* > *sp*. Can change secondarily to *šp* (e.g., in NP *šipiš* ‘louse’, via assimilation) ~ *sb* 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 *\*žy* > *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  $*h_u > f$  (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  $*fr > \check{s}$  (also seen in Balochi).

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  $*-r_0- > / \rightarrow -u-$  in the present stem of ‘do’ (Sogdian *kwn-*, Yaghnobi *kun-* — 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 Σαρμάται ‘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  $*r_i > l$  (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 ταπαγγάς, cited in Tzetzes’ *Theogony*, Oss I *dæ bon xorz*, D *dæ bon x<sup>w</sup>arz* (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  $*śu$ ,  $*žy$ . He describes an Old Central Iranian “represented by most of the remaining dialects” that has undergone the changes  $*śu > sp$ ,  $*žy > zb$ . 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  $*ś > s$ ,  $*ž > z$  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 requireOrmuri 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., *arót* ‘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  $\gamma$ - 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 *\*ɣ-* (e.g., Oss *uad* ‘wind’ < PIr *\*uāta-*).<sup>6</sup> Khotanese Saka has *b-* [β] and *d-* [ð], 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 *\*ɣ-* (as well as *\*β-* and *\*ð-* in the case of Tumshuqese). Yaghnobi shows *d-*, plausibly from *\*ð-* (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,

<sup>6</sup>Fragments of Scythian, thought to be closely related to Ossetic, show an orthographic merger between reflexes of PIr *\*ɣ-* and *\*b-*, e.g., *\*baiɣar-ašɣa-* > Βαίρασπος ‘(having) myriad horses’, *\*ɣanat-ašɣa-* > Βανάδασπος ‘(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’ < *\*gauza-* (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 PIIr *\*-ft-*, *\*-xt-*

East Iranian languages tend to voice PIIr *\*-ft-*, *\*-xt-*; West Iranian languages tend to devoice PIIr *\*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). Édel’man (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:

$$T+t \rightarrow T^{(s)}t \qquad D+t \rightarrow T^{(s)}t \qquad D^h+t \rightarrow D^{(z)}d^h$$

Old Avestan preserves BL across the board,<sup>7</sup> e.g., *aogədā* ‘utter’ 3sg inj. (< PIIr *\*Haugd<sup>h</sup>a* ← *\*\*Haug<sup>h</sup>-ta*); this is generally undone in Young Avestan, e.g., *aoxta* ‘id.’ (with some exceptions, e.g., *vərəzda-* ‘grown’ : OInd *vṛddhá-*, *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 *D<sup>h</sup>* (in the case of PIIr *-b<sup>h</sup>-*, they would have to metanalytically rely on the pattern shown by *-p-*, since no roots ending in PIE *\*-b-* survived in PIIr). Édel’man argues the following:

1. PIIr *\*b<sup>h</sup>*, *\*d<sup>h</sup>*, *\*g<sup>h</sup>* had fricative allophones [β], [ð], [ɣ] (p. 53) which were preserved in East Iranian and Avestan (keeping a distinction between PIIr *D* and *D<sup>h</sup>*).
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-*; Édel’man 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

<sup>7</sup>Though Rix et al. 2001:658, fnn. 4–5 argues that the *f* of OAv *vaf-* ‘sing’ goes back to a participle *\*u(a)fta-* ← *\*\*u(a)b<sup>h</sup>-ta* in which BL was undone; additionally, if OAv *duuafša-* ‘torment’ is to be connected to Greek *tuphlós* ‘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 (D<sup>h</sup>-s → T-s on the basis of T-s, D-s), e.g., PIIr *\*gr<sup>h</sup>b<sup>h</sup>-sca-* → *\*grf-sa-* > 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 *ubj-* ‘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 *βt*, *γt* 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 *fəδ*, *xδ* (cf. Monna 1978, *passim*)). Yaghnobi devoiced 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-* > *\*βaytam* > MSog *βxtm* /*vaxtam*/ ‘*devātideva*’ (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 Gandhārī Prakrit fully assimilates heterorganic clusters, e.g. *satamma* ‘7th’ < OInd *saptamā-*, 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-ǎfta-*. 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 *\*-u<sup>h</sup>fta-* < *\*-u<sup>h</sup>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-iabda-*; 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 *aqil-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 *\*kaup<sup>h</sup>a-* (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-b<sup>h</sup>-*, 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., *kaofō* nom. sg. (Yt 19.3), *kaofanəm* 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., *\*madakā-* vs. *\*madaxā-* ‘locust’, *\*sūrākā-* vs. *\*sūrāxā-* ‘hole’, *\*āiaka-* vs. *\*āiāxa-* ‘egg’, *\*ahmāka-* vs. *\*ahmāxa-*, etc. (Klingenschmitt 2000:203, fn. 40), so parallels would be welcome.

It seems likely that PIIr *\*kaup<sup>h</sup>a-* adheres to the same sort of derivational pattern as PIE *\*roth<sub>2</sub>o-* ‘thing with wheels, chariot’ ← *\*rot-eh<sub>2</sub>-* ‘wheel’, e.g., *\*koup-eh<sub>2</sub>-* ‘hump’ → *\*kouph<sub>2</sub>o-* ‘thing with a hump/peak/dome’.<sup>8</sup> 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 *\*maiθā-* ‘day’** PIr *\*maiθā-* ‘day’ > Shg *mēθ*, Yzg *miθ*, Zebākī *mī*, Ygh *met*, Yidyā-Munjī *miš*, Sangličī *mēi*, Orm *mīθ* (prob. a Shughni loan), Buddhist and Manichean

<sup>8</sup>Thanks to Michael Weiss for this suggestion.

Sogdian *myδ*, Christian Sogdian *myθ* and Khwarezmian *myθ* /me:θ/. Morgenstierne (1938:229) argues against the diphthong \*-ai-, since Sariqolī has the monophthong *māθ* (cf. *mēyz* ~ *māyz* ~ *mayz* ‘urine’ < \**maiža-* (Paxalina 1971:267)), and instead sets up a protoform \**māθia-*. He gives references connecting the form to Lithuanian *mėtas*, Albanian *mot* ‘time’, a cognate of Old Indic *māti-* ‘measure’ (Demiraj 1997:278). PIr \**māθ(i)ia-* would be a thematized derivative of an abstract noun \**māti-*. It is not clear if this form works for other languages, e.g., Shg *xūbaθ* ‘him/herself’ < \**hūai-paθia-* (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 *zyrn* ‘gold’ < \**zārania-*, *pyrδnn* ‘saddle’ < \**pāri-dāna-* (Sims-Williams 1989e:181), but it’s not clear that this account explains the shape of the other forms. If we assume an original \**māθ(i)ia-*, metathesis to \**maiθa-* must have occurred in Khwarezmian, for which PIr \**θi* > *s*, e.g., \**hūai-paθia-ka-* > *xbsk* ‘own’ (Humbach 1989:195).

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

Many of these languages seem to preserve the timespan sense, e.g., Buddhist Sogdian *’str’ myδ* ‘later in the day, the next day’ (Gershevitch 1954:13), Yidyā *do yū miḥ* ‘in one day’ Morgenstierne (1938:loc. cit.) (but Srq *maθ 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 *maēθa-*, and has developed the semantically natural sense of ‘day’, possibly independently, in each of these languages. Sariqolī 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 *kup*, Psht *kab* ‘fish’; Modern Ossetic has *kæsaɡ*, 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 *iæu-gæf* ‘caviar, roe’ (lit. “fish millet”). Cf. also the Scythian hydronym Παντικάπης < \**panti-kapa-*, lit. ‘fish path’; a connection between these forms and Elamite *Ka<sub>4</sub>-ab-ba* 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 *kapanā-*, Latvian *kāpe* ‘caterpillar’, noting the putative cognacy between Slavic \**ryba* (< \**rūbā-*) ‘fish’ and German *Raupe* ‘caterpillar’ (however, the connection with *kapanā-* 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*, *rūpe* ‘caterpillar; eelpout’ exhibits polysemy indicating a natural semantic connection

<sup>9</sup>The following step is unnecessary, but worth mentioning: if we accept Insler’s (1971:174) reading of OAv *maēθā* ‘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 \**siōk* ‘lodge the night, pass a night’ (Matisoff 2003:323, 328).

between CATERPILLAR and (ELONGATED) FISH.<sup>10</sup> 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(yn)* ‘to heat’ (Thordarson 2009:154). Cheung (2002:18–9) claims that PIr *\*p* “occasionally” yields postvocalic final *-f*, picturing a development *\*-pV̄ > \*-bV̄ > \*-v > -f*.<sup>11</sup>

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 *kavaṃñe*, *kaviñe*, *kaviñāṃ*) from *\*kapā-*. from *\*kapā-*; the plural in *-e* 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 *\*hižuan-* (against other Iranian *\*hižu(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 *\*hižuH(ā)-* ‘tongue’.<sup>12</sup> Some Iranian forms are unsuffixed, e.g., *\*hižūā-* > YAv *hizuuā-*, Psht *žəba* (*\*-aka-* should become Psht *-ay*), Shg *ziv*. We also see the *\*-aka-* suffix across most of East Iranian, a widespread Indo-Iranian feature: PIr *\*hižūāka-* > Sogdian *zβ’k*, C *zb’q*, Khwarezmian *zβ’k*, Khotanese Saka *bišāa-* /βiʒa:/, Yzg *zveg*, Ygh *zivok*, Wkh *zik*, Zēbakī *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. Parachi (*zu*)*bān* and Ormuri *zubān* are marked as loans by Morgenstierne (1929:302, 413). All West Iranian words for ‘tongue’ reflect PIr *\*hižūāna-*, e.g., OP *h-z-a-n-m* (acc.sg.), MP *’zw’n* /uzwa:n/, 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 *ziwān*, Maz *zevon*. Given this distribution, it is possible to reconstruct a Proto-West Iranian form *\*hižūāna-*, 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 *\*hižuan-* 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 *\*dafānam* ‘mouth’ acc. sg. (cf. YAv *zafar-*

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

<sup>11</sup>Both *kæf* (which he derives from *\*kapā-*) and *kæræf* ‘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. *æxsæv*, D. *æxsævæ* ‘night’, with *-v-*, it remains unclear how an *-an-* stem could yield the proper conditioning environment for final devoicing of *\*-v*.

<sup>12</sup>On the irregular historical phonology of this form, see Fay 1895:ccxxviii; Lipp 2009:I 188–90.

‘mouth’, acc. sg. *θrizafanəm* ‘having three mouths’), *\*dantānam* ‘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 *lingua* (classical *lingua*) ‘tongue’ is an *ā*-stem, but the Germanic cognate exemplified by Gothic *tuggō* 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 *āh-* ‘mouth’ (Y 28.11 *ēāāṇhā* inst. sg., Y 31.3 *āṇhō* gen. sg.), Young Avestan has *āṇhan-* ‘mouth’ nt. (V 3.29 *āṇhānō* gen.sg.). In RV, we see *ās-* alongside *āsān-*; it is not clear, however, whether the *an*-stem is reconstructible to PIr. In Young Avestan, we also see *vīmītō.dantan-* ‘having deformed teeth’ (V 2.29 *vīmītō.dantānō* nom. pl.).

It seems like the most likely locus of the extension of *an*-stems is PIr *\*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 *āṇhānō* gen.sg. and Ved *āsāni* loc.sg.). This inflectional pattern could then have been extended to YAv *dantan-*, and also to *\*hizūan-* in some Old Iranian dialects. Forms like heteroclitic YAv *zafar/n-* ‘mouth’ along with derivatives like *θrizafana-* could also have provided influence. Ultimately, while *\*hizūāna-* 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 Mazandarani (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 $*\theta$

Old Persian and Median preserve Proto-Iranian  $*\theta$  (though in some contexts on Old Persian, it is palatalized). However, intervocalically,  $*\theta > h$  in most West Iranian languages (Balochi undergoes fortition to  $t$ ).<sup>13</sup> This is in sharp contrast to East Iranian, where  $*\theta$  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  $*\theta$  that was different from its counterpart in Proto-East Iranian? Or perhaps this is the wrong question to ask, given the diachronic instability of  $[\theta]$ . Was a similar factor at play in East Iranian, or was there some sort of sociolinguistic pressure responsible for the longstanding preservation of  $*\theta$  in these languages?

An additional confound is the fact that  $*\theta n$  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 $*\theta n$

Southwest Iranian has  $\check{s}n$  for  $*\theta n$ :

- $*d\bar{a}\theta na-$  > Aramaic  $d\check{s}n$ , Elamite  $da\check{s}-na$  = OP  $*d\bar{a}\check{s}na-$  ‘gift’, with a long vowel, according to Tavernier (2007:407)<sup>14</sup>
- $*ara\theta ni-$  > OP  $ara\check{s}ni-$  ‘cubit’ > Phl  $’l\check{s}n$  MMP  $’ry\check{s}n$  /a:reʃn/ > NP  $\check{a}re\check{s}(n)$
- $*dm\bar{a}na-pa\theta n\bar{i}-$  > MMP  $b’nby\check{s}n$  /ba:mbiʃn/
- $*ham-pa\theta n\bar{i}-$  > NP  $\bar{a}mvasn\bar{i}$  (?) ‘rival wife’

We also see Phl  $’lnc$  > NP  $\bar{a}ranj$  ‘elbow’, perhaps from a source closely related to Sogdian (cf.  $’rync$ ).

The clearest evidence of the outcome of  $*\theta n$  in NW Iranian is seen in the Babylonian Akkadian transmission of a Median PN: Akk  $Pa-at-ni-e-\check{s}a$  = Med  $*Pa\theta n\bar{i}\check{e}\check{s}a-$  <  $*Pa\theta n\bar{i}-ai\check{s}a-$  ‘looking for [desiring?] a wife’ (Tavernier 2007:273). The  $t$  points unambiguously to  $*\theta$ , indicating that the cluster was unchanged in Old NW Iranian. Gk  $\Pi\iota\sigma\sigma\acute{o}\upsilon\theta\nu\eta\varsigma$  =  $*Pi\check{s}i\check{a}\theta na-$

<sup>13</sup>Proper names like *Mithradates* and demonyms like *Part’ew* are generally taken as evidence that “an older stage of Parthian” (Korn 2005:81) preserved  $*\theta$  (Sundermann 1989:123).

<sup>14</sup>Schwartz (1974:401) argues that MMP  $d’\check{s}(y)n$  /da:ʃin/, Phl  $d’\check{s}n$  (as well as NP  $d\bar{a}\check{s}an$ ?) cannot reflect  $*d\bar{a}\theta na-$  due to spelling, and because MPth  $d’\check{s}(y)n$  cannot be borrowed from MP. He argues for an etymology  $d\bar{a}\check{s}in\bar{a}-$  ‘that which is presented or indicated with the right hand’, cf. OInd  $dak\check{s}i\check{n}a-$  ‘right (handed)’ →  $dak\check{s}i\check{n}\bar{a}-$  ‘gift due to priest (for performance of a rite)’. Cf. also Bct  $\lambda\alpha\beta\nu\omicron$  ‘gift’ (Sims-Williams 2007:226).

(with haplology, cf. YAv *Piši.šīiaovna-*, a bahuvrihi with the second element meaning ‘actions’ and the first unknown,<sup>15</sup> 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’š(y)n* = MMP *d’š(y)n* (though perhaps not a reflex of *\*dāθna-*), *b’nbysn* = MMP *b’nbysn*. 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’man (2003:I 213) give Bal *hariš*, *harš*, *harša* and Awr *aražni* ‘cubit’ (*\*-šn-* > Awr *žn*). 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., *\*čašman-* > *cam(m)* ‘eye’. Benveniste (1935:105) derives the Zazaki *-iš* infinitive suffix *-iš* from *\*-θna-* (with *\*n* lost, and a preceding thematic vowel).

A number of Iranian loans into Armenian show the change *\*-θn-* > *n*. Gippert (1993:348) identifies Armenian *danak* ‘knife’ as a borrowing of NW Middle Iranian *\*dahnak* < PIR *\*daθna-ka-* (alongside Arm *dašnak* ‘id.’ < SW Middle Iranian *\*dašnak* < PIR *\*daθna-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 *\*axšaina-* ‘brown’ or *\*xšaiθn̄-*, taking *\*axšēn* to be a possible MP outcome (perhaps via dissimilation?). He also cites Oss *äxsin* ‘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 *\*θn* was unchanged, but most Middle and New NWIr languages show *šn* (or a secondary change from it, e.g., Awr *žn*, Bal *n* ~ *š[?]*). Some of this may be due to Persian lexical borrowing. But in the case of the Zazaki infinitive suffix (if < *\*-θna-*), borrowing seems unlikely; why would Zazaki borrow the (albeit deverbal) MP abstract noun suffix *-išn* 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.

## 2.6.3 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.

### 2.6.3.1 Reflexes of PIR *\*š*, etc.

Ilya Gershevitch (1962) argued that *\*š*, *\*ž* > OP *θ*, *d* was not a legitimate sound law and instead represented some sort of register variation.

<sup>15</sup>Perhaps related to OInd *piš-* ‘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,  $\sigma\acute{\rho}\acute{\alpha}\chi\alpha$  ‘dog’, recorded by Herodotus (1.110.1). Given this item, Gershevitch suggests, we can be reasonably confident that words with *sp* for PIr  $*\acute{s}\acute{u}$  are of Median origin (e.g., *aspa-* ‘horse’ alongside *asa-bāra-* ‘horseman’), but we cannot be equally confident that variation of the type  $*\acute{s} > s \sim \theta$ ,  $*\acute{s}\acute{i} > \acute{s}y \sim \theta y$  (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  $*ga\acute{z}na-$  ‘treasure’ (> Pth,Sogd  $\gamma zn$ , NP *gašn* ‘galore’), found indirectly in Achaemenid-era documents as  $*ganza-$  and  $*ganda$  (e.g., Elamite *kan-da-ba-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  $*\acute{z}n$  (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  $*\acute{z} > d$  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  $*\acute{z} > d$  was only an optional, register-specific rule.<sup>16</sup>

Gershevitch (1962:19–22) takes  $\theta y \sim \acute{s}y$  alternation as evidence for variation between  $\theta$  and *s* as reflexes of PIr  $*\acute{s}$  in Old Persian. Others (cf. Hoffmann 1976b:637, fn. 25) disagree; Klingenschmitt (2000:203) ascribes alternation between pre-OP  $*-i\acute{i}a-$  and  $*-i\acute{a}-$  to analogical suffix alternation (“Suffixwechsel”), not phonological conditioning. Pre-OP  $*\theta\acute{i}$  would yield  $\acute{s}iy$ , and pre-OP  $*\theta\acute{i}i$  would yield  $\theta iy$ . This accounts for some variation within Persian:  $*t\acute{u}\theta\acute{i}i\acute{a}k\acute{a}- > MP tuh\acute{g}$ ,  $*m\acute{a}\theta\acute{i}i\acute{a}-k\acute{a}h\acute{i}a > MP m\acute{a}h\acute{g}$ , vs.  $*k\acute{a}\acute{s}i\acute{a}p\acute{a}-k\acute{a}- > MP ka\acute{s}avag$  (expected †*kašabag*), NP *kašav*, *kašaf*,  $*\acute{u}a\acute{s}i\acute{a}h- > MP wyš$  /we:ʃ/ > NP *bēš* ‘more’. This leaves variation in onomastic items like Elamite *Ši-ia-ma* = OP  $*Sy\acute{a}va-$ <sup>17</sup> versus Elamite *Ti-ia-ma* = OP  $*\theta y\acute{a}va-$  ‘black’, if we are willing to identify them with  $*\acute{s}i\acute{a}ua-$  (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  $*\theta\acute{i} > *s\acute{i}$  precedes consonant + glide epenthesis. Word-initially, consonant + glide epenthesis appears to pre-date  $*\theta\acute{i} > *s\acute{i}$  change, since MP *sy’h* and NP *syāh* ‘black’ can reflect either OP  $*\theta\acute{i}i\acute{a}ua-$  or  $*s\acute{i}i\acute{a}ua-$ , but not  $*s\acute{i}i\acute{a}ua-$  (this is my main reason for taking the OP onomastic item  $*\theta y\acute{a}va-$  seriously). For this reason, it looks as though pre-Old Persian contained a doublet  $*\acute{s}i\acute{a}ua- \sim *s\acute{i}i\acute{a}ua-$ , the disyllabic member of which underwent palatalization (which was blocked in the trisyllabic member). Given that this  $*-i\acute{i}a- \sim *-i\acute{a}-$  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  $*\acute{s}\acute{i}$  and  $*\theta\acute{i}$  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

<sup>16</sup>Cf. 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).

<sup>17</sup>Elamite *š* also = OP *s*; however, see Asatryan 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  $h < *θ$  in their word for ‘fish’, except for Parthian *m’sy’g* /ma:sja:g/, Zazaki *māse* and Kurdish *māsi*. PIr *\*š(i)īāua-* shows *siy-*, e.g., MP *sy’h* > NP *syāh* ‘black’, Zaz *siyā*, Qohrūdī *sigāh*, with the characteristic loanword phonology seen in other Persian borrowings (e.g., NP *piyāz* → Qoh *piġāz* ‘onion’). Furthermore, PIr *\*kašīapa-* shows a plain *s* in more than one language, e.g., Bandarī *kāsapošt* ‘turtle’,<sup>18</sup> Balochi *kasīp* ~ *kāsib* ‘turtle, tortoise’. (Korn 2005:284 notes that a “genuine” Balochi word should contain *š*. However, long *ī* reflects disyllabic *\*iīā*; cf. Korn 2005:105. It could then be the case that *kasīp* reflects *\*kàšīāpakà(hīa-)* gen.sg.).<sup>19</sup> 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 *\*šū*. 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  $σπάα$ , 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* ~ *sf* for PIr *\*šū* 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. *zū* (and *sū*) should necessarily become either *zb* (*sp*) or *z* Khot. *ž* (*s* Khot. *ś* Wkh *š*) and exclude the possibility of other developments?...The modern Persian *sf=sp* seems to be confined to words with original *sū*...(barring a few arabicized forms) and could be attributed to the influence of a dialect in which *sū* resulted directly in *sf*.”

New Persian shows at least three relatively consistent outcomes (i.e., limited to particular etymological reflexes) for PIr *\*šū*:

1. *\*šū* > *s*; e.g., *\*ašūa-bāra-* > OP *asabāra* ‘horseman’ > Phl *’sw’r* /aswa:r/ > NP *suwār* ‘rider’; *\*šūaka-* > NP *sag* ‘dog’
2. *\*šū* > *sp* ~ *šp*; e.g., *\*ašūa-* > NP *asb* [æsp] ‘horse’, *\*šūiš-* > NP *sepeš* ~ *šepeš* ‘louse’
3. *\*šū* > *sf* ~ *sp*; e.g., *\*šūanta-* > NP *esfand* ~ *espan* ‘wild rue’; *\*šūaita-* > NP *sefid* ~ *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* ~ *šepeš* ‘louse’ (< Phl *spyš*) does not have an attested antecedent in Old Persian,

<sup>18</sup>It is possible that this form reflects an anticipatory dissimilation *\*š...š* > *s...š*, rather than *s < \*šī*. However, assimilation of the type [s ... š] > [š ... š] 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.

<sup>19</sup>In a sense, *pace* Korn, it may be the case that *sy* and *šy* are not, respectively, Persian and Balochi outcomes of these clusters. Multiple languages show these outcomes depending on whether *\*š* 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 *wsp*, MMP *wysp* < OP *vispa-*. But forms in which *sp* and *sf* vary may come from another source. Martin Schwartz (p.c.) remarks that it is unlikely that *sf* came about via a secondary change from *sp*, given that instances of *sp* from etymological sources other than PIr *\*śu* do not undergo such a development. Schwartz (2006:223) says that for *espan* ~ *esfand* ‘wild rue’, the variant with *sf* “cannot be due to the mediation of Arabic, which has only *ħarmal*.”

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, *Gabrī* (a derogatory term; NP *gabr* ‘non-Muslim’ = Turkish *gavur* = Arabic *kāfir* ‘infidel’), stands alongside a Jewish lect. Some Zoroastrian Yazdi forms seem to have *sv* for PIr *\*śu*, e.g., *s(e)vīd* ‘white’ (Bailey 1936:349), *svaka* ‘dog’ (Krahnke 1976:231, citing the field notes of Michael M.J. Fischer; Bailey gives *sag*). From the terms given by Bailey, it is hard to tell whether *sv* continues a preform *\*su* or *\*sf*—his Yazdi’s *v* reflects (along with *\*m* and *\*u*) *\*f*, e.g., *\*drafša-* > *dravš* ‘banner’—or if it could possibly come from “Median” *sp*. According to Vahman and Asatrian (2002:20), PIr *\*śu* “remains unchanged in a very important lexeme: **sva**, **seba**, **sawa**, **sewa** ‘dog’.” However, the presence of the word *svarz* ‘spleen’ < PIr *\*spr̥za(n)*—seems to indicate that PIr *\*sp* also became *sv* in Yazdi. So while it is likely that NP forms with *sf* are loanwords, Yazdi is unable to explain their distribution (i.e., why *sf* for *\*śu*, but not *\*sp*?) — the issue raised above by Henning.

### 2.6.3.2 PIr *\*r* + coronal clusters

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

- PIr *\*zrd-* > MMP *dyl* /*dil*/ > NP *dil* ‘heart’
- PIr *\*yarda-* > NP *gul* ‘flower’
- PIr *\*šarda-* > Phl *šNT* (Aram.) /*sa:l*/ (MMP *s’r*) > NP *sāl*
- PIr *\*p(a)rdanku-* > NP *palang* ‘panther’ (cf. Vedic *pṛdāku-* ‘snake’, meaning ‘tiger, panther’ in lexica, SSog *purδ’nk* ‘panther, leopard’)<sup>20</sup>
- PIr *\*br̥zant-* > MMP *bulnd* /*buland*/ > NP *buland* ‘high’
- PIr *\*barzād(a)*-<sup>21</sup> > MMP *b’l’y* /*ba:la:j*/ > NP *bālā* ‘height’
- PIr *\*marz-* > Phl *m’l-* /*ma:l-*/ ‘rub, sweep’ > *mālīdan* ‘rub, polish’
- PIr *\*r̥zif̥a-* > MMP *’lwp/f* /*a:luf*/ > NP *ālūh* ‘eagle’

<sup>20</sup>This form is likely a Wanderwort; see Lubotsky 2001.

<sup>21</sup>Perhaps a de-instrumental *d*-stem built to PIE *\*b<sup>h</sup>erg<sup>h</sup>-eh<sub>1</sub>-*, cf. Latin *mercēs*, *mercēdis* ‘wages’, *herēs*, *herēdis* ‘heir’ (Weiss 2009:304-5).

- PIr *\*zarnu-mani-* ‘gold neck’ > NP *dāl-man* ‘black eagle’ (Schwartz 1971:292, fn. 14)<sup>22</sup>
- PIr *g(a)rna-ka-* > NP *galla* ‘flock’ (Schwartz 1971:292, fn. 14)
- PIr *\*prtū-* → *\*prθū-* > MMP *puhl* /puhl/ > NP *pūl* ‘bridge’

This change, from what we can see, post-dates Old Persian.<sup>23</sup> 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 *sipil* ‘id.’ < PIr *\*spržan-*. On the outcome of PIr *\*rd/rž* 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/t* as native or loan-words their preponderance is significant.”

We encounter other vexing aspects of the distribution of *l*. For instance, Gorani *zil* ‘heart’ shows a decidedly non-SW Iranian outcome of PIr *\*ž-*, 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 *d*? 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 *\*wi-nāša-* > MP *wināh* > NP *gunāh* ‘sin’
- PIr *\*wi-čāra-* > Phl *wc’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 *gwlbk’* /gurbag/ > NP *gurba* ‘cat’ (cf. YAv *urupi-* ‘dog, fox (?)’ < *\*urpi-*)
- PIr *\*u(i)jāna-* > MP *gyān* > NP *jān* ‘life, soul’
- PIr *\*uahāna-ka-* > MP *wihān(ag)* > NP *bahāna* ‘reason. pretext’ (cf. Digor Ossetic *ræuonæ* < PIr *\*fra-uahāka-*, from *\*uah-* ‘dress’; see Gershevitch 1952:483-4)

<sup>22</sup>Assuming that this form is a word equation with YAv *zarənumainiš*; *dāf*<sup>o</sup> on its own could also be connected with PIr *\*zarta-* (assuming that YAv *zairita-* ‘golden’ ← *\*zarta-* due to the influence of *zairi-* ‘gold’).

<sup>23</sup>There are no good direct precursors of the forms listed above, but we see forms like OP *ardata-* ‘silver’ < PIr *\*(a)ržata-*, cf. Yazdi *ālī* (Kent 1951:171).

- PIr *\*uī-dāna-* (?) > Phl *wyd'n*, MMP *wy'n /wijan/* > NP *giyān* ‘tent’ (cf. OInd *vi-dhā-* ‘furnish, spread, diffuse’?)
- PIr *\*uata-* > Phl *wʹ*, MMP *wd /wad/* > NP *bad* ‘bad’
- PIr *\*uāta-* > Phl *wʹt*, MMP *wʹd* > NP *bād* ‘wind’
- PIr *\*uī(n)šati-* > Phl/MMP *wyst /wi:st/* > NP *bīst*
- PIr *\*u(a)rda-*<sup>24</sup> > Phl *gwl /gul/* > NP *gul* ‘flower, rose’
- PIr *\*uarka-* > Phl *wlg /warg/* > NP *barg* ‘leaf’
- PIr *\*uahia-* > MP *wah* > NP *bah-*
- PIr *\*uašia-* > MP *wyš /we:ʃ/* > NP *bēš* ‘more’
- PIr *\*uahišta-* ‘best’ > Phl *whšt*, MMP *whyšt /wahift/* > NP *bihīšt* (cf. the name of a 4th cent. CE Christian martyr, *Gwšt'z'd*, the first member of which < *\*uahišta-*)
- PIr *\*uarna-ka-* > Phl *wlkʹ*, MMP *wrg /war:ag/* > NP *barra* ‘lamb’
- PIr *\*uart-* ‘turn’ > OP *v-r-t-* > Phl *wlt-*, MMP *wrd-* /ward/ > NP *gard(ūn)* ‘wheel, chariot’ (Cheung 2007:424–5)
- PIr *\*uažr(a)-ka-* > OP *v-z-r-k /vazrka-/*<sup>25</sup> > Phl *wc(w)lg*, MMP *wzrg /wuzurg/* (cf. Pazand *guzurg*, Bailey 1933:56) > NP *buzurg*
- PIr *\*uažra-* > Phl *wlz /warz/* > NP *gurz* ‘mace’ (cf. Bal *burz* ‘club’, Elfenbein 1963:25)
- PIr *\*ua/injēča-ka-* (?) > Phl *wncškʹ*, Psalter *wnčšky /windʒiʃk/* > *binjišk* ~ *gunjišk* ‘sparrow’, cf. Baxtiyārī *bingišt* (Schapka 1972:236);
- PIr *\*ur̄injī-* > Phl *blnc /brindʒ/* > NP *birinj* ~ *gurinj* ‘rice’ (cf. AV+ *vr̄ihī-*)

The historical phonology of the forms given above shows some dominant trends: we see that PIr *\*ur-* generally > MP *gur-*, PIr *\*uī-* > MP *wi-* usually > NP *gu-*. Elsewhere, we see PIr *\*u-* > MP *w-* > Early New (Judeo-)Persian [β-]<sup>26</sup> > NP *b-*, but there are some environments where it is difficult to make a generalization, particularly the fate of *\*uar-*, which shows change to both *g-* and *b-* in conditioning environments that are not particularly clear, as well

<sup>24</sup>YAv *varəda-* ‘rose’ points to (and Semn *val(a)* ‘flower’ seems to point to — perhaps also Pth *wʹr war/*) *\*uarda-*, while the Persian forms, along with Gor *wilī*, may point to *\*urda-* (MacKenzie 1961:77 gives the former etymon for all these forms).

<sup>25</sup>Schmitt (1989a:69) and others give this reading — departing from *vazraka-* (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.

<sup>26</sup>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  $*\underset{\cdot}{u}$ -, e.g., *vazay* ‘frog’, *vazīdan* ‘blow’, *varzīdan* ‘work’, *vīr* ‘memory’ (MacKenzie 2003:110); the verbs are marked as loans by Cheung (2007:426, 431).

The forms descended from  $*\underset{\cdot}{u}a\underset{\cdot}{z}r(a)$ -(*ka*-) were likely NWIr loans into Old Persian since they show *zr* as opposed to expected *dr* (though our best evidence of the change  $*\underset{\cdot}{z}r > OP\ dr$  is word-initial, e.g., *drayah-* ‘sea’). In any case, the relevant historical phonology (i.e., changes to  $*\underset{\cdot}{u}$ -) 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  $*\underset{\cdot}{u}$ -  $> b$ - the regular SWIr outcome. MacKenzie (1971a:76) takes the change  $*\underset{\cdot}{u}$ -  $> b$ - 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  $*\underset{\cdot}{u}$ - is often modified, forms with *g-* and *b-* preponderate, likely due to contact. For instance, Kumzārī *ward* ‘flower’ sits alongside *bīsta* ‘twenty’ (Thomas 1930:814, 825).<sup>27</sup> Additionally, while Zazaki usually shows *v-* (e.g., *vā* ‘wind’), the word for ‘blood’ is *gūnī*  $< *\underset{\cdot}{u}ahuni$ -<sup>28</sup> (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.

<sup>27</sup>Though Martin Schwartz (p.c.) notes that this form, rather than being tantalizingly conservative, is likely a loan from Arabic.

<sup>28</sup>The word for blood shows irregular historical phonology across west Iranian. NP *xūn* (MP *xōn*) has either undergone a metathesis between  $*\underset{\cdot}{u}$  and  $*h$ , or was subject to the same irregular *x*-prothesis as MP *xāyag*, NP *xāya* ‘egg’. Parthian has *guxn*, with unexpected *g*-. Sivandi has *fīn*.

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 discontinuous 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 (*\*iažata-* '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 morphosyntactic 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 Wāziri; I added features from Mazandarani 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’ (< \*č̣juta-), 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 \*θ** Old Iranian *θ* 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 *θ* is seen in Khwarezmian: PIr \**čaθuāra-* > Pth, Khw *cf’r* /*cafar*/ ‘four’. No New East Iranian language retains *θ* as such, except Šuyūnī (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 **theráï** alongside **sériáï** 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 *δry*, Manichean Sog *šy* ‘three’.

**3.2.1.1.3 PIr \*šr > 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 *niyaçāray-* ‘lean’ caus. (× *dāraya-* ‘hold’ caus., Kent 1951:188), MP *ns’y* /*nisa:j*/ ‘conveying, dispatch’. Elsewhere, Middle Persian and New Persian preserve the cluster \**šr*, e.g., \**šrāuaia-* ‘hear’ (caus.) > MP *sr’y* /*sara:j*/ ‘sing’, NP *sarāy*.

New Persian shows a number of doublets:

- PIr \**ḥyašrū-* > NP *xusrū*, *xušū* ‘mother-in-law’

- PIr *\*h<sub>2</sub>uaśura-* > NP *xasar*, *xusū* (perhaps with elision of final *r* rather than a continuation of OP *\*ç*) ‘father-in-law’
- PIr *\*aśra(-ka)-* > NP *ašk*, *ars* ‘tear’ (cf. Khw *šwk?*)

It is possible to view *xusrū* 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 *xusū* 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 Šuynī *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’; *\*h<sub>2</sub>uaśru-* > Balochi *vassū* ‘mother-in-law’.

In short, the outcome of PIr *\*śr* 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 *\*ś<sub>u</sub>* > *sp*** The development of the Proto-Iranian cluster *\*ś<sub>u</sub>* is a strong diagnostic of subgrouping. The change PIr *\*ś<sub>u</sub>* > *s* defines SW Iranian, and the change PIr *\*ś<sub>u</sub>* > *š/śś* unites Khotanese, Tumshuqese, and Wakhi. However, the rest of Iranian has *sp* or something incredibly similar. The change *\*ś<sub>u</sub>* > *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’ (< *\*ś<sub>u</sub>aka-*) given by Herodotus (1.110), as well as the Scythian proper name *Βαιορασπος* (< *\*bai<sub>u</sub>ar-aś<sub>u</sub>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.<sup>1</sup>

**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.

<sup>1</sup>Interestingly, 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, where we see *sp* in some words, and Ossetic, where we see *s* < *\*š* in one word ‘flea’. YAv *pasuuqm* ‘cow’ gen pl is a clear restoration of the stem *pasu-*. It is not clear how to account for Yazdi *svaka* (cited by Krahnke 1976:231 from the field notes of Michael M.J. Fischer)—is it an incredible archaism, just one step removed from PIr *\*ś<sub>u</sub>*, or a secondary change from *\*sp* or *\*sf*? Bailey (1936:347) cites a Yazdi form *s(e)vid* ‘white’, but gives *sag* for ‘dog’.

**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, Ōrmuṛī, Parāčī, Iškāšmī 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  $*ā$ ;  $*fš > kš$ ;  $*śr > ś$ ;  $*š > ʃ$ ;  $*θi > thh$ ;  $*-r...n- > -r...ṇ-$ , but  $*rn > rr$ . In Pashto, Wakhi, and Iškāšmī,  $*śr > ʃ$ . In Pashto and Iškāšmī,  $*rś > ʃt$ . In Ōrmuṛī,  $*rś > ʃ$ . In Pashto, Iškāšmī, and Parāčī,  $*rt$ ,  $*rd > r$ . In Pashto,  $š$  in Persian loans is retracted to  $ʃ$ . Some Parāčī dialects undergo a change  $rz > rʒ$ .

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 Parāčī, 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 Parāčī). 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  $-tæ$ , 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  $-āre$  (Schwartz 1969).

**3.2.1.2.3 Causative in  $*-āyāia-$**  Causatives and denominatives in  $*-āyāia-$  are identified for Khwarezmian and Khotanese Saka in Schwartz 1969. Present also in the Wakhi causative suffix  $-uv-$ .

**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, Kumzārī, Baškārdī, Lārestānī, Kurdish, Zāzākī, Semnani, Gorani, and Central Dialects, the  $ezafe$  is head marking. In Zāzākī, the clitic inflects for gender. The following example shows a New Persian noun phrase with an  $ezafe$ .

(3.1)  $gol =e sorx$   
 flower =EZ red  
 ‘red flower’

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

- (3.2) *isbi* =*a* *asb*  
 white =EZ horse  
 ‘white horse’

This innovation postdates the Old Iranian period. The development of the *ezafe* can be traced from Old Persian to Middle Persian via the grammaticalization of the relative pronoun *haya-* into an enclitic. Constructions like *Gaumata haya maguš* ‘Gaumata the Magian’ (Darius, Behistūn, *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 *mē-* imperfective** While the *ezafe* postdates Old Iranian, the *me-* 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 Mazandarani; Kurdish, Balochi, and Zazaki have other means of marking imperfective verbal forms.

**3.2.1.2.6 Preverbal *bi-/be-* subjunctive/perfective** The preverb *bi-* 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, Zāzākī, and Balochi. In Mazandarani, it retains its perfective-aspect-marking role.

**3.2.1.2.7 Object marking with *rā*** Middle Persian *rāy*, grammaticalized from OP *rādiy* ‘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, Ābyāne’i...Bandar Abbasi, Minabi, Baškardi, Kumzārī, Balōčī 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 *hnd* ‘blind’, Zaz *aqil-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 *kæsaɡ*, which Thordarson (2009:58) connects to Waṇetsi *kəžə* ‘*id.*’, replacing an older form *kæf* (< *\*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 *xratu-* and Old Persian *xraθu-* ‘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*, Lārestānī *mard*, but Kumzārī *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 *\*d<sup>h</sup>ǵ<sup>h</sup>om-* ‘earth’ vs. *\*d<sup>h</sup>ǵ<sup>h</sup>emon-* ‘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<sup>2</sup> (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 *zər* ‘*id.*’, but both reflect PIr *\*žrd-*). To this data I added Zazaki (Paul 1998b), Kumzari (Thomas 1930), Baškārdī (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 Yidghā-Munjī, 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 *\*ž* (here, the etymon is *\*žrajah-*) only in SW Iranian (and only part of the time!).<sup>3</sup>

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 *našq* ‘ornament’ < Tajik Persian *naqš* (Édel’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,

---

<sup>2</sup><http://ielex.mpi.nl/>

<sup>3</sup>The word *\*žasta-* ‘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 Hasmik Kirakosian which mentioned an instance of *zast* in the Old Āzarī *Fahlavīyāt*, 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-tāpa-*) appears in Pamir languages, e.g., Shg *ōftōb*.<sup>4</sup>

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 *mūyi* ‘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 *\*spržan-*. We expect PIr *\*rž* > 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/rž* 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/l̥* as native or loan-words their preponderance is significant.”) The Swadesh-200 word list does not contain ‘spleen’ (though it contains ‘liver’, Yagnobi *šupurda*<sup>5</sup>), 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 *\*spržan-* 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 *qenë* ‘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 *\*hižūā-* or *\*hižūāka-*, with *\*hižūāna-* 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 *\*hižūā(ka)-*. So the change of state due to the loan may not be highly relevant.

---

<sup>4</sup>Though the evidence is somewhat murky, we generally expect the *\*-b-* of *\*abi-* to be voiced in Pamir languages, e.g., Yazghulāmī *avden* ‘bridle’ < *\*abi-dān(d)ia-*, *avdūst* ‘mitten’ < *\*abi-dasta-* (Édel’man 1980:303), Šuyūnī *biyis-* ‘swell, inflate’ < *\*abi-haič-(sa-)*, *bidāj-* ‘irrigate’ < *\*abi-tač-* (Cheung 2007:127).

<sup>5</sup>This form shows the nature of complications posed by loanwords—this form has the distinctly Yagnobi trait of palatalizing earlier *\*sp* < PIr *\*śu* (seen also in Ishkashmi and Pashto), but has the distinctly Persian trait of showing *d* for PIr *\*ž*. Fortunately here, Yagnobi 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 *’eškāl* ‘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* /e:k/ (< *\*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 *\*rāšta-* ‘right’ > *rāst*** Kent (1951:34) considers a change of PIr *\*rāšta-* → OP *rāsta-* ‘straight, right’ to be due to analogy with past participles of dental stems (PIE *\*-t<sup>s</sup>-t-* > PIr *\*-st-*). The OP form stands against Avestan *rāšta-* 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 *ōftōb* ‘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-* intervocally, 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

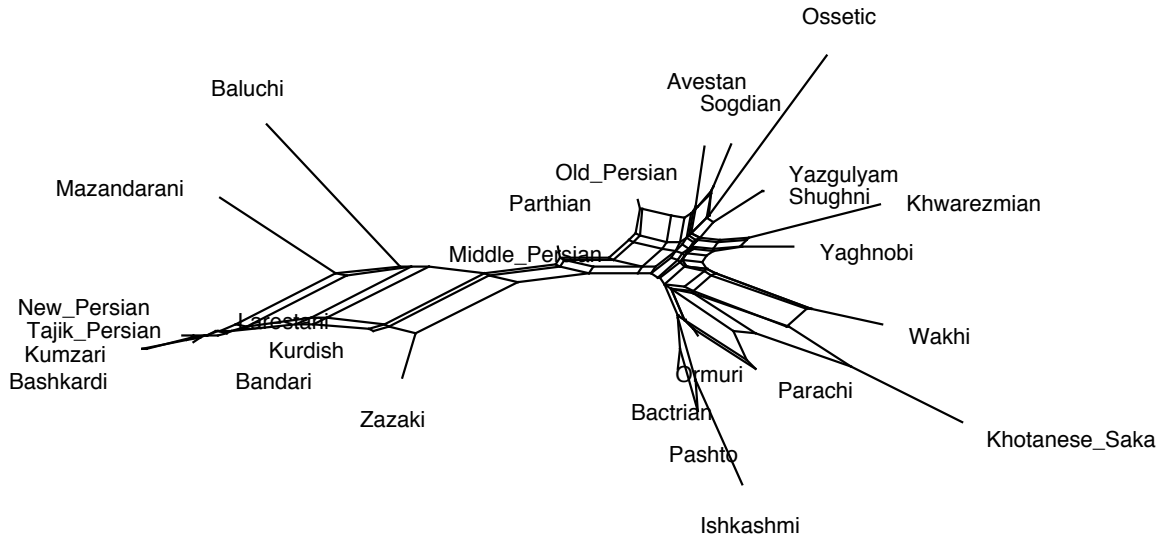
This section presents the results of simulations based on character types discussed in the previous section.

### 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.

Figure 3.1: NeighborNet with Binary Typological/Recurrent Characters



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.)

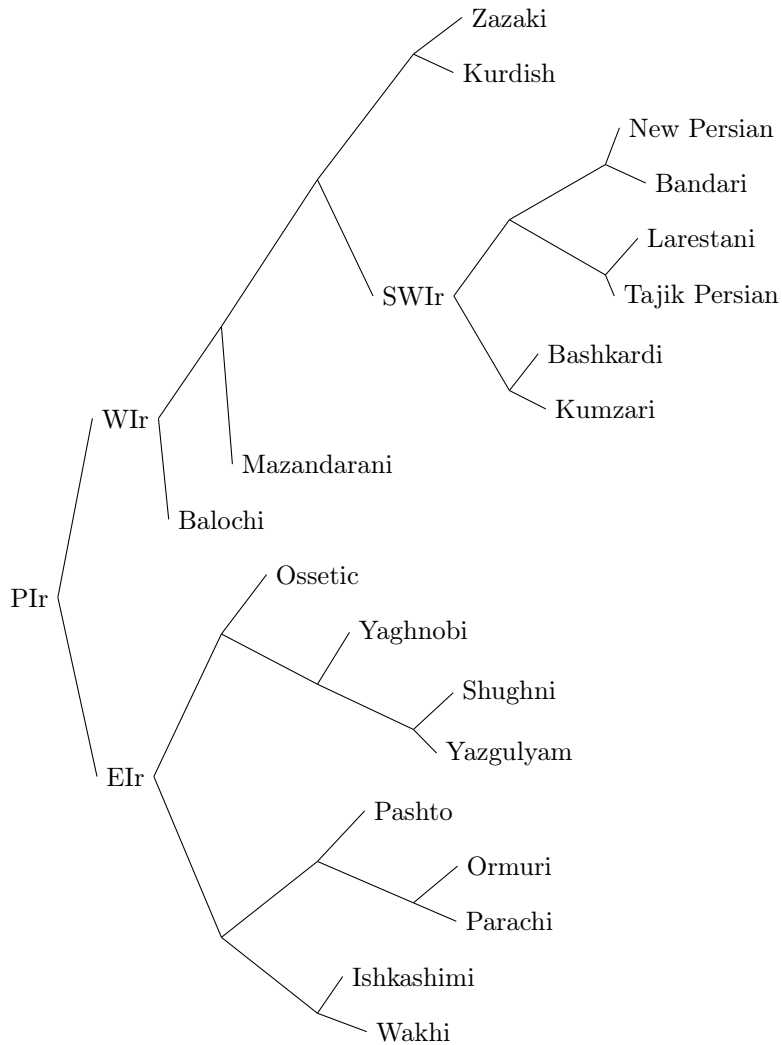
The summary tree for the non-lexical data is shown in Figure 3.2. Clear East and West Iranian clades are visible and annotated.

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

Figure 3.2: Typological data; consensus tree



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 SWIranian clade within a West Iranian clade is produced, but no NWIranian clade. No East Iranian clade is produced.

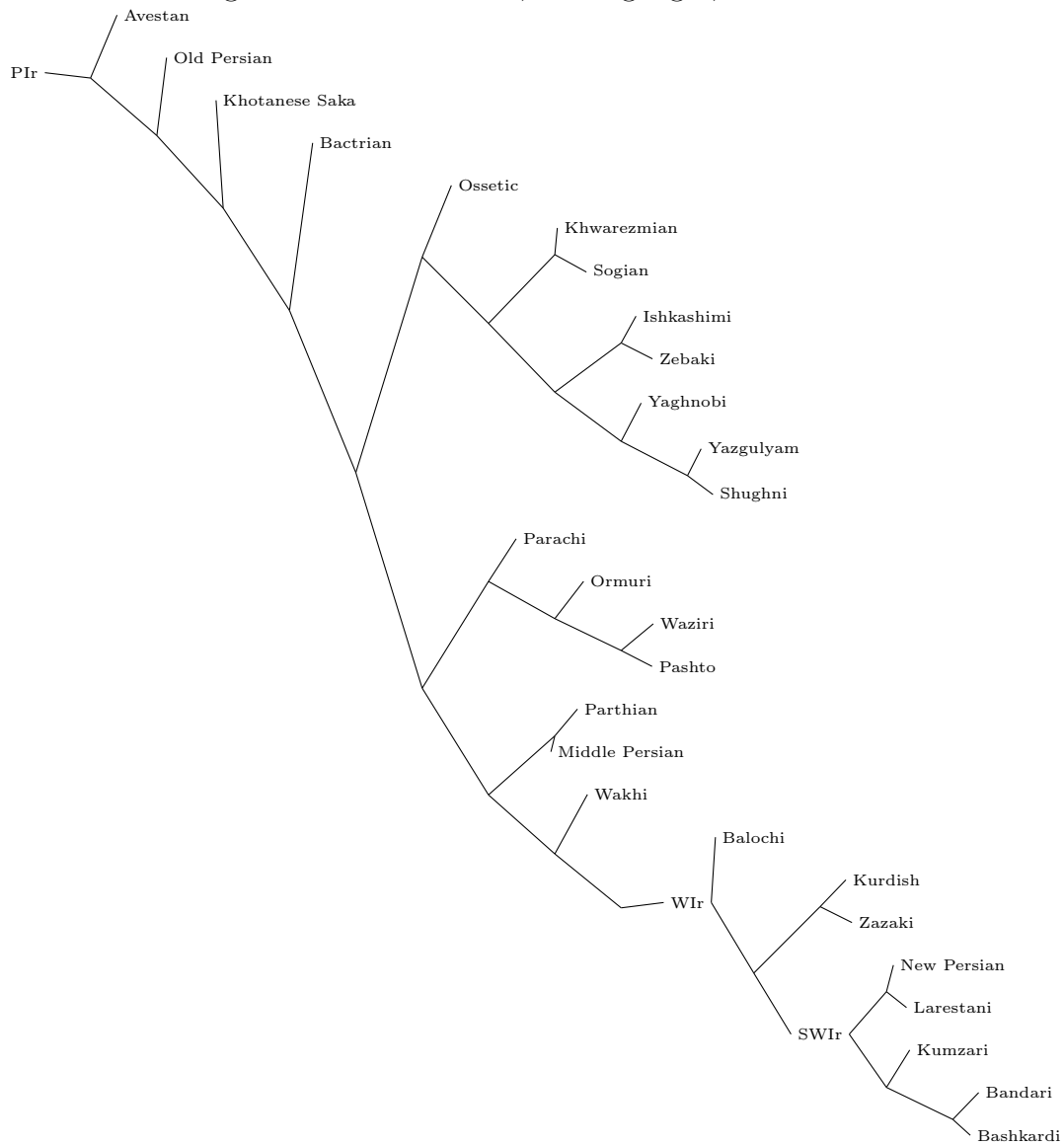
**3.3.2.2.2 Loans uncoded** The summary tree for the dataset of only contemporary languages in which loans went uncoded is shown in Figure 3.6.

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.

Figure 3.3: Lexical data; all languages; loans coded



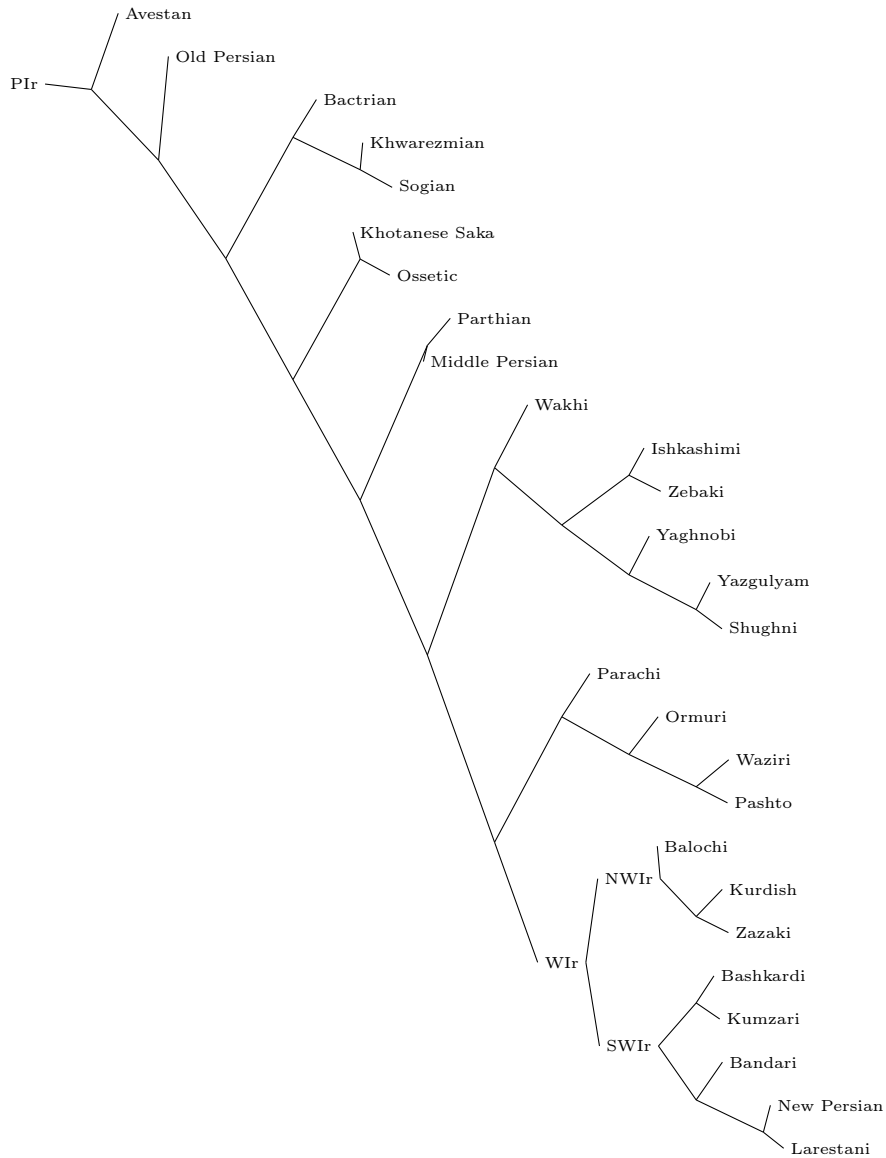
### 3.3.2.3 Summary

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

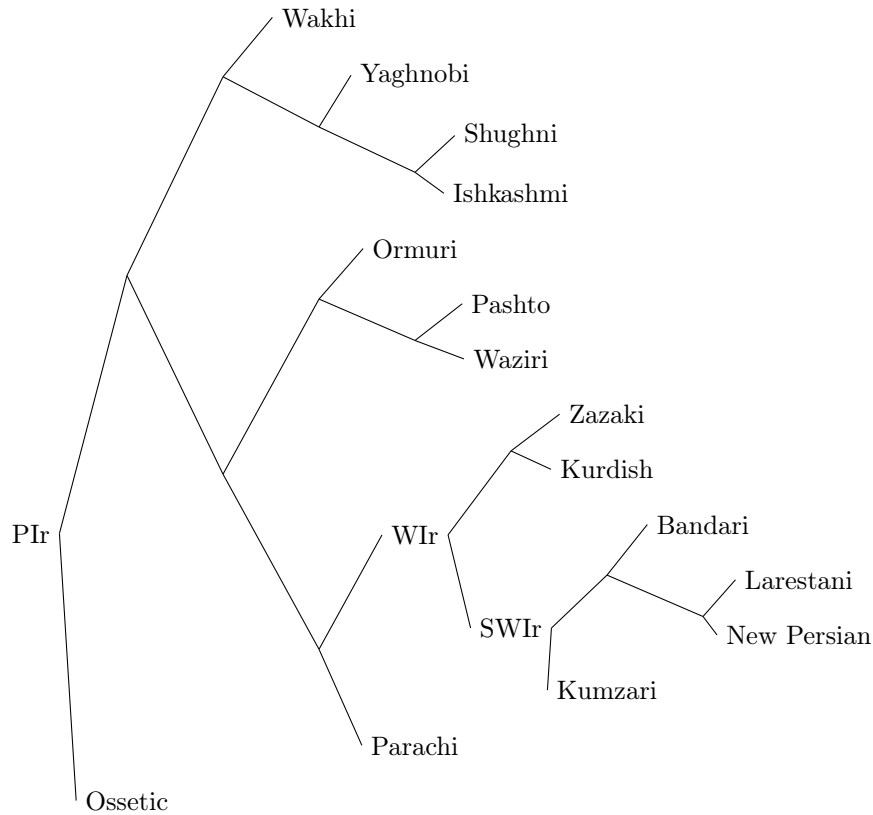


Figure 3.4: Lexical data; all languages; loans uncoded



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

Figure 3.5: Lexical data; contemporary languages; loans coded



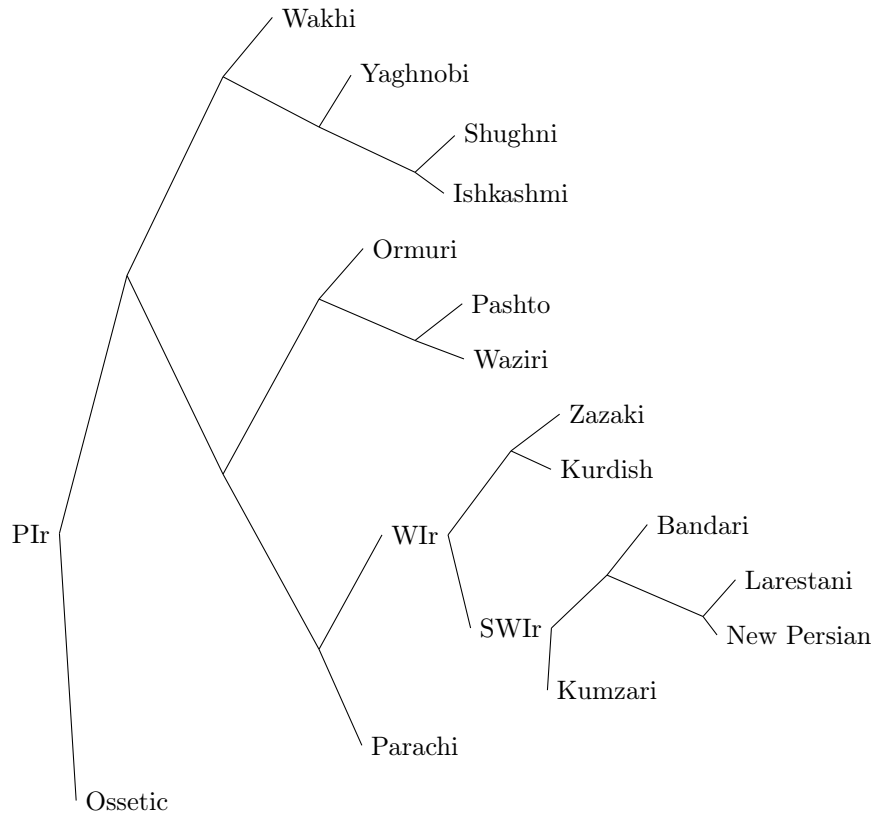
loans. Unfortunately, the lack of a documented historical record for languages other than Persian make it difficult to tell whether a West Iranian subgroup is supported by lexical data at all points throughout history, and is not in fact a product of advergence or prolonged contact.

### 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

Figure 3.6: Lexical data; contemporary languages; loans uncoded

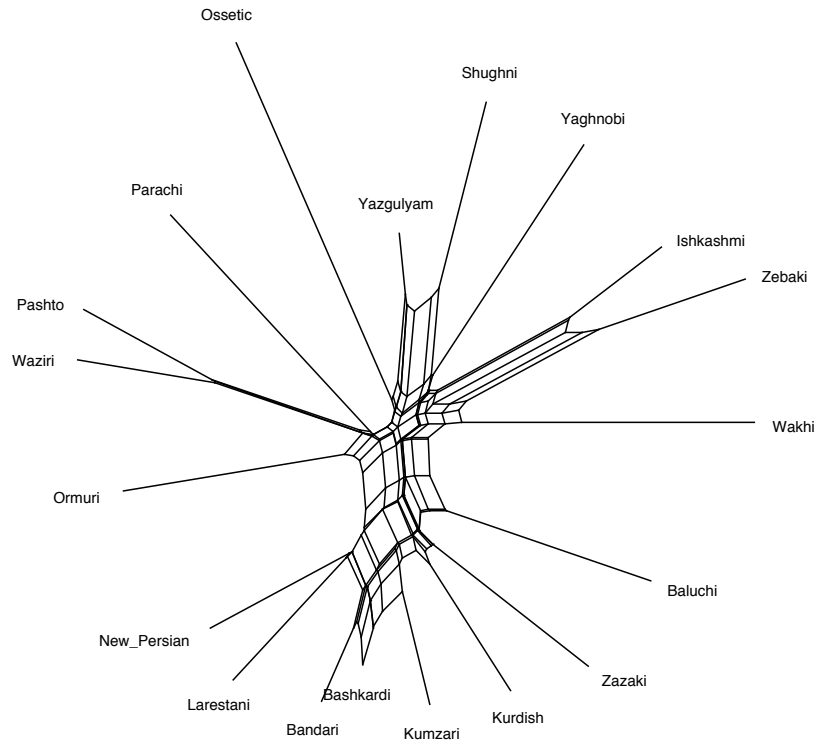


3. Distance-based analyses of lexical character data produces an East-West division with no clear split strength
4. Distance-based analyses of typological and other non-lexical recurrent characters produces an East-West division, albeit with a great deal of webbing around the center of the network

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

Figure 3.7: NeighborNet with Lexical Characters, loans coded



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  $\theta$ ), 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., *\*hižyan-* ‘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

1

### 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

---

<sup>1</sup>This 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-grammarians hypothesis yields a significant result.”

This sort of appraisal of irregular-looking change is not so easy to carry out for the functional items under study in this chapter. A functional item in a given non-Persian West Iranian language may be identical to its Persian counterpart and may appear to have undergone an irregular sound change, but this irregular change may not be a clearly identifiable regular Persian development. Furthermore, irregular sound change (i.e., irregular phonological reduction and analogical change) is quite common in grammatical morphemes. Hence, it is difficult to determine whether the irregular change in the non-Persian language indicates a borrowing, or an actual irregular sound change. More broadly, it is not clear if large-scale formal congruence among West Iranian functional items is due to borrowing (from Persian, most likely), or shared tendencies expected among genetically related languages.

To address this question, this chapter employs a quantitative methodology designed to measure amounts of irregular sound change in a subset of a language’s functional vocabulary. In order to investigate the degrees of irregular sound change affecting forms in West Iranian languages, I analyze etymologically related functional items in two Middle and eight New West Iranian languages, some of which appear to have undergone various sorts of irregular phonological reduction and analogical change in the course of their development, in the manner described below.

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* (Raxšānī 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), *Awromani* (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



Figure 4.1: Approximate geographic locations of West Iranian languages treated in this chapter



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 *\*ž* 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 *\*žrd-* 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 *\*rādi*) 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 *\*rādi* to become Bal †*-rād*. 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 *nāwih̄t* ⇒ Modern English *nought*; it can involve elision of segments, or both; it can involve gestural overlap, e.g., Modern English *did you* ~ [ˈdɪdʒə] ~ [dʒʌ]; and it can involve combinations of these phenomena, e.g., Basque *ez dakit* ‘I don’t know’ ~ [stait] (Trask 2000:15).<sup>2</sup>

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 cognates<sup>3</sup> in other West Iranian languages (if they exist; I deal with complications

---

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

<sup>3</sup>I 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.

stemming from missing data by using mixed-effects linear regression, as described in §4.5).<sup>4</sup> 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.<sup>5</sup> 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 *rād* 'on account of, for the sake of, because of, for', Bal *-rā* indirect/direct object marker, Maz *-rə* IO/DO marker, Zaz *-rē*, *-rā* IO marker, Sang *-r(e)* IO marker, Yzd *-ra* oblique marker, Awr *-ra* IO marker

2. NP *hič* 'any, nothing,' ('aucun, rien, en rien', Lazard 1957:119-20)

- PIr *\*aiya-čī-*
- OP *\*aiya-čī-* 'one thing' > *\*ēwč* (with syncope, cf. Skjærvø 2009a:201) > MP *ēc* 'any' ⇒ NP *hič* 'anything' (with Cockney-style prothesis)
- Pth *ēwiž* 'any', Bal *hičč-ī* 'something', Kd *hič* 'any', Maz *hečč(ī)* 'nothing', Sang *hauču* 'nothing', Yzd *heš(ki)* 'nobody, none', Awr *hič* 'any'

3. NP *har* 'each, every'

- PIr *\*harua-*
- OP *haruwa* 'all' > MP *harw* 'all, every' > NP *har* 'each, every'

<sup>4</sup>I 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 *\*iā-* (relative pronoun) → OP *haya-* ⇒ MP *-ī(g)* (*ezafe*) > NP *-i*. However, we do not know whether other West Iranian *ezafes* (if not simply borrowed from Persian) developed from a relative pronoun *\*iā-*, 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 *\*iā-*, 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)').

<sup>5</sup>All but the attested Old Persian transcriptions have been normalized (OP *c* = *č*, OP *iy* = *ī*, 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 *\*hačā* ‘in association with’
- OP *hacā* ‘from’ ⇒ MP *az* > NP *az*
- Pth *až*, Bal *aš* (Makrani *ač*), Kd *ž*, Maz *-jā*, Sang *az*, Yzd *az*, Awr *ja*

5. NP *ki* (complementizer)

- PIr *\*kahja*
- OP *kahya* ‘what’ gen. sg. > MP *kē* ‘who, which’ ⇒ NP *ki* (complementizer)
- Pth *kē* ‘who, which’, Bal *ki* (complementizer), Kd *ka*, Maz *ke*, Zaz *ki*, Sang *ku*, Yzd *gī*, Awr *ka*

6. NP *agar* ‘if’

- PIr *\*ha-kr̥(-)t-* ‘once’
- Pre-OP *\*ha-kr̥(-)t-* → OP *hakaram* (on the basis of *kara-* ‘maker’, Kent 1951:212) ⇒ MP *agar* ‘if’ > NP *agar*
- Pth *ag*, Bal *agar*, Kd *agar*, Maz *agər*, 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)<sup>6</sup> > 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 *\*dūtīa-ka(ra)-*
- OP *\*dūtīakàram* ‘2nd time’ ⇒ MP *dudīgar* ‘2nd’ ⇒ NP *dīgar* ⇒ *digar* ‘other’
- Pth *bidīg* ‘2nd, other’, Bal *digar* ‘other’, Maz *dəgər*, Sang *dīgar*

9. NP *mē-* (imperfective prefix)

---

<sup>6</sup>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.).

- PIr *\*ham-aiua-*
- OP *\*hamaiwa-* > MP *hamēw* ‘always’ ⇒ NP *mē-* imperfective prefix
- Pth *hamēw* ‘always’, Sang *mī* (imperfective prefix), Awr *ma-/mə-*

10. NP *magar* (adversative connective)

- MP *ma agar* ‘hopefully not’ ⇒ NP *magar* ‘but’
- Bal *magar*, Maz *magə*, Yzd *mager*, Kd *magar* ‘unless, if not’

11. NP *yak* ‘one’<sup>7</sup>

- PIr *\*aiua-ka-*
- OP *\*aiuakahia-* gen. sg. (*aiva-* attested) > *\*ēiákē* (adapted from Cantera 2009:19) ⇒ MP *yak*<sup>8</sup> (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, yoā*

12. NP *injā* ‘here’

- MP *ēn gyāg* ‘this place’ (⇐ virtual *\*(h)aima-* [a demonstrative not yet present in Old Persian] + OP *\*wiyāka-*) ⇒ NP *injā* ‘here’
- Maz *īnje*, Zaz *injā*, Sang *anjū*, Awr *injā*, Kd *injā* ‘then, in that case’

13. NP *dar* ‘into’

- PIr *\*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 *dē* ‘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 PIr *\*h-* in *\*hačā*, but regular, language-specific changes to PIr *\*-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.

<sup>7</sup>While not a functional item *per se* (and not the primary means of marking indefiniteness in most West Iranian languages), this item is generally of high frequency and prone to contraction (cf. also poetic Hindi/Urdu *ek* ~ *ik* ‘one’).

<sup>8</sup>This 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 *-rā* (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 †*-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 /dɔg/ and /fɪɔg/ 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 *\*b<sup>h</sup>ud<sup>h</sup>-(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  $x^1, \dots, x^n$ . A response value  $y_i$  is modeled as a linear function of predictor values  $x_i^1, \dots, x_i^n$ , plus a value drawn from a normally distributed error term  $\epsilon \sim N(0, \sigma_\epsilon^2)$ . The formula can be written as  $y_i = \beta^0 + \beta^1 x_i^1 + \dots + \beta^n x_i^n + \epsilon_i$ , where  $\beta^0$  denotes the intercept and  $\beta^1, \dots, \beta^n$  denote the slopes associated with the predictor vectors  $x^1, \dots, x^n$ . 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  $y_i = \beta^0 + \beta^1 x_i^1 + \dots + \beta^n x_i^n + \epsilon_i$  also characterizes a regression with one  $n$ -level categorical predictor variable. For level  $k = 1, \dots, n$ ,  $x_i^k = 1$  and  $x_i^{-k} = 0$ . 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.,  $u_j^0 \sim N(0, \sigma_u^2)$ .

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  $y_{ij} = \mu^{\text{NP}} + u_j + \vec{\beta}^{\text{Lang}} \cdot X_i^{\text{Lang}} + \epsilon_{ij}$  for each functional item  $j = 1, \dots, J$  in language  $i \in \{\text{Balochi, Kurdish, Mazandarani, Zazaki, Sangesari, Yazdi, Awromani}\}$ , where  $\mu^{\text{NP}}$  is the intercept;  $u_j$  is a deflection for each functional item  $j = 1, \dots, J$ ;  $\vec{\beta}^{\text{Lang}}$  is a 7-length vector of language-specific slopes for Balochi, Kurdish, Mazandarani, Zazaki, Sangesari, Yazdi, and Awromani;  $X_i^{\text{Lang}}$  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 `lme4` (Bates et al. 2013) with the call



function `lmer(distance ~ Balochi + Kurdish + Mazandarani + Zazaki + Sangesari + Yazdi + Awromani + (1|functional_item), data)` (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 DEGREE OF IRREGULAR PHONOLOGICAL CHANGE (i.e., the Levenshtein distance between the expected and observed forms) as a function of (New West Iranian<sup>9</sup>) LANGUAGE, with random intercepts by 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 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  $\mu^{\text{NP}}$  in the formula given in the previous section), but represents language-specific estimated slopes for the remainder of the languages.

The `dropterm()` function from R’s 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).

---

<sup>9</sup>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

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

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

Lang.	$\hat{\beta}_{LD}$	$\hat{\beta}_{RLD}$	$\hat{\beta}_{LLD}$	$\hat{\beta}_{QC}$	$SE_{LD}$	$SE_{RLD}$	$SE_{LLD}$	$SE_{QC}$	
MP	0	0	0	0	NA	NA	NA	NA	$N = 10$
Pth	0	0	0	0	NA	NA	NA	NA	
NP	1.625	0.221	0.881	1.125	0.399	0.045	0.147	0.309	$N = 14$
Bal	0.875	0.138	0.315	1.5	0.226	0.041	0.095	0.267	
NP	1.625	0.251	0.895	1.375	0.386	0.056	0.164	NA	$N = 14$
Kd	0.25	0.02	0.063	0	0.25	0.02	0.063	NA	
NP	1.5	0.308	0.824	1.25	0.677	0.126	0.308	0.568	$N = 8$
Maz	0.5	0.083	0.1277	0.25	0.5	0.083	0.127	0.25	
NP	1	0.2	0.693	0.5	NA	NA	NA	0.79	$N = 4$
Zaz	0	0	0	0.5	NA	NA	NA	0.5	
NP	1.666	0.361	0.963	1.666	0.235	0.065	0.095	0.471	$N = 6$
Sng	0.333	0.083	0.135	0.333	0.333	0.083	0.135	0.333	
NP	1.571	0.275	0.866	1.285	0.368	0.06	0.17	0.28	$N = 14$
Yzd	-0.142	-0.028	-0.057	0	0.26	0.061	0.105	0.21	
NP	1.5	0.285	0.895	1.5	0.27	0.076	0.109	0.27	$N = 8$
Awr	0.25	0.035	0.101	0.25	0.25	0.035	0.101	0.25	

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)<sup>10</sup> 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.

---

<sup>10</sup>This 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.

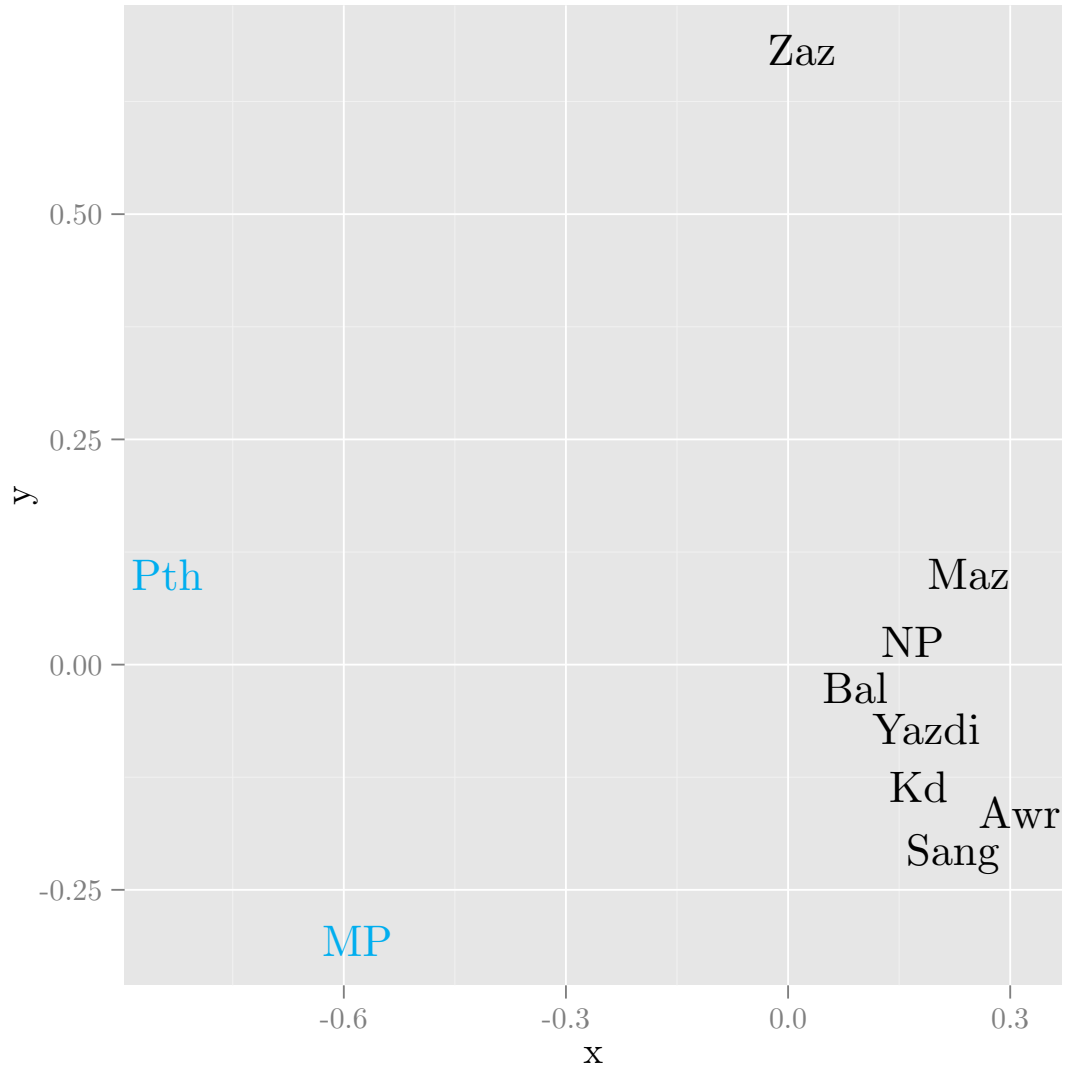


Figure 4.2: Multidimensional scaling of inter-language distances

Table 4.3: Distance between Proto-Iranian etyma and lexical items in different languages

Lang.	$\hat{\beta}_{LD}$	$\hat{\beta}_{RLD}$	$\hat{\beta}_{LLD}$	$SE_{LD}$	$SE_{RLD}$	$SE_{LLD}$
NP	2.84	0.59	1.28	0.18	0.03	0.04
Bal	-0.44	-0.09	-0.13	0.15	0.03	0.04
Kd	0.28	0.03	0.09	0.3	0.06	0.08
Maz	0.24	0.03	0.07	0.29	0.05	0.08
Zaz	0.47	0.06	0.10	0.29	0.05	0.08
Sang	0.32	0.06	0.11	0.3	0.06	0.08
Yzd	0.62	0.12	0.16	0.3	0.06	0.08
Awr	-0.006	-0.006	0.01	0.31	0.06	0.08

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  $\mathbf{u}$  and  $\mathbf{v}$  is the total number of all vector indices  $i$  where  $\mathbf{u}_i \neq \mathbf{v}_i$ . 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 *\*āraθni-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 *\*ž* (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 *\*āja-(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 *\*rž*).

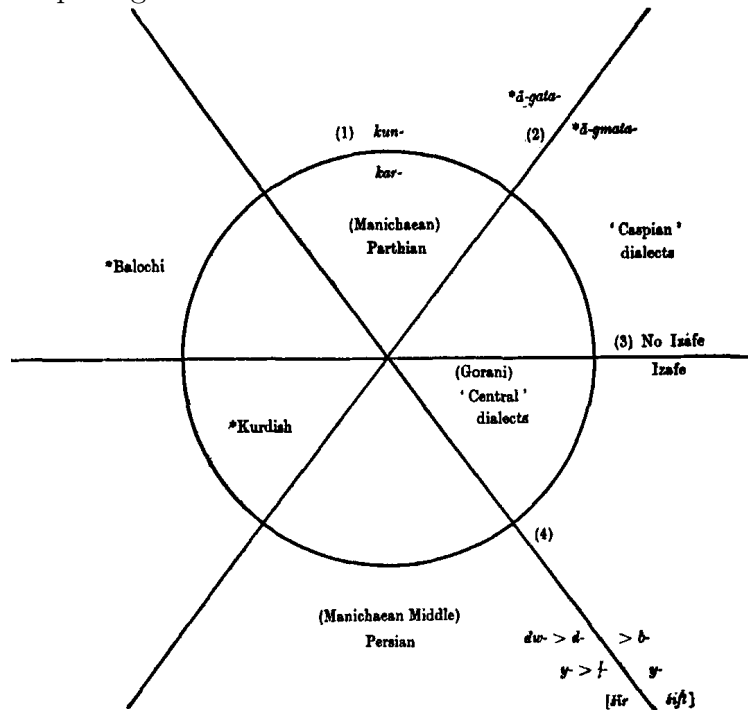
This approach has yielded some valuable observations on the nature of contact between West Iranian languages, but can be somewhat atomistic. A desideratum in Iranian dialectology is a model that allows us to look at agreement of this sort across the lexicon, not just in individual lexical items.

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., *\*š* and *\*ž*, as well as *\*ṣ̌u* and *\*ẓ̌u*, *\*ṣ̌i*, *\*ṣ̌r*, etc.). We are relatively certain that reflexes like *h* < *θ* < *\*š*, *s* < *\*ṣ̌* are Southwest Iranian in origin (though recall that Gershevitch 1962 believes that only *s* < *\*ṣ̌* is a secure SWIr reflex). The provenance of chronologically shallower changes is less clear. For example, the clusters *\*rž*, *\*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 *\*u̇-* > *b-* ~ *g-* ~ *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 *\*u̇-*, 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., *\*š* > *h* for NP, despite the fact that it often

shows *s*).

Figure 5.1: Schema of West Iranian dialects in MacKenzie 1961:71, according to non-varying phonological and morphological features



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. PIr *\*laupāša-* ‘fox’
  - (a) *h* IN REFLEX OF *\*laupāša-* (e.g., NP *rubāh*)
  - (b) *s* IN REFLEX OF *\*laupāša-* (e.g., S Tātī *luās*)
2. PIr *\*spržan-* ‘spleen’
  - (c) *rz* IN REFLEX OF *\*spržan-* (e.g., NP *supurz*)
  - (d) *l* IN REFLEX OF *\*spržan-* (e.g., Kd *sipil*)
3. PIr *\*uāta-* ‘wind’

- (a) *b-* IN REFLEX OF *\*uāta-* (e.g., NP *bād*)
- (b) *g-* IN REFLEX OF *\*uāta-* (e.g., Bal *gwāt*)
- (c) *v-* IN REFLEX OF *\*uāta-* (e.g., S Tātī *vār*)

In a given language, etymological doublets are coded, even if a variant is clearly due to contact (e.g., Baločī has both *zird* and *dīl* 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 Ėdel’man 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 PIR *\*žarnu-* or *\*žar(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 in 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:

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

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,  $*\acute{s}u$  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  $\alpha = .91$ ). 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  $*\acute{s}$ ,  $*\acute{s}i$ ,  $*\acute{s}u$  ( $N = 15$ )
  - LATERAL DISTANCE: features pertaining to PIr etyma containing  $*r/r_t$ ,  $*r/r_d$ ,  $*r/r_z$  ( $N = 26$ )
  - 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

	Great-circle distance		Least-cost distance	
Overall linguistic distance	$R^2 = .16$	$p = .002^{**}$	$R^2 = .11$	$p = .01^*$
Palatal distance	$R^2 = .10$	$p = .01^*$	$R^2 = .03$	$p = .17$
Lateral distance	$R^2 = .29$	$p < .001^{***}$	$R^2 = .21$	$p < .001^{***}$
Labiodental distance	$R^2 = .03$	$p = .19$	$R^2 = .03$	$p = .17$

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,  $.16 \leq R^2 \leq .38$ . 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 .16, but  $p$ -values below .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. Goebel (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  $*r/\underset{\cdot}{r}t$ ,  $*r/\underset{\cdot}{r}d$ ,  $*r/\underset{\cdot}{r}z$ ) correlates significantly ( $p < .001$ ) with great-circle distance and least-cost distance

2. Palatal distance (i.e., distance based on reflexes of  $*\acute{s}$ ,  $*\acute{s}i$ ,  $*\acute{s}u$ ) correlates significantly with great-circle distance ( $p = .01$ ) and insignificantly with least-cost distance
3. Labiovelar distance (i.e., distance based on reflexes of  $*\acute{u}$ -) correlates insignificantly with both great-circle distance and least-cost distance

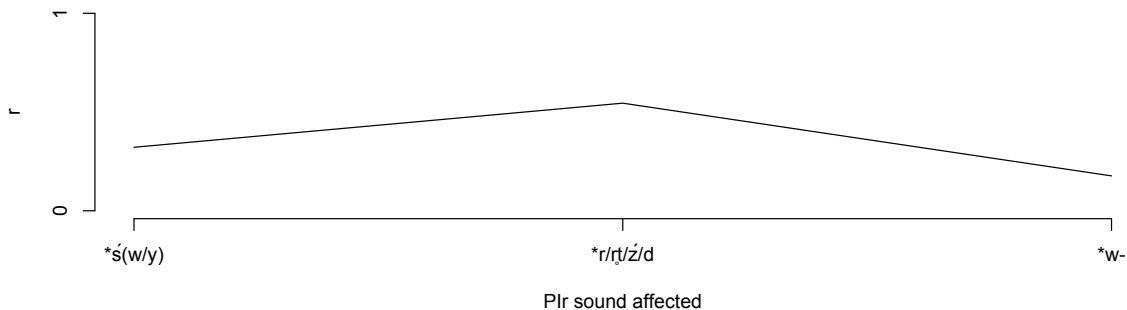
Do these results reflect the intuition that recently actuated and diffused features correlate more strongly with geographic distance than more older ones (keeping in mind the caveat that lexical items containing features that spread at an earlier date can be borrowed at any time)?

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

- The change of PIr  $*\acute{s}u$  to  $s \sim sp$  predates Old Persian, e.g., PIr  $*a\acute{s}ua-$  > OP  $asa-$  ~  $aspa-$  (< Median?) ‘horse’
- The change of PIr  $*r/\acute{r}t$ ,  $*r/\acute{r}d$ ,  $*r/\acute{r}z$  to  $l$  postdates Old Persian and predates Middle Persian, e.g., OP  $b(a)rd-$  ‘be high’ > MP  $bul(ana)$  ‘high’
- The change of PIr  $*\acute{u}$ - to  $b-$  postdates Middle Persian and predates New Persian, e.g., MP  $wad$  > NP  $bad$  ‘bad’
  - The dating of the change  $*\acute{u}$  >  $g-$  is less clear; PIr  $*\acute{u}i-nā\acute{s}a-$  > MP  $wināh$  > NP  $gunāh$  ‘sin’, but PIr  $*\acute{u}r\acute{t}ka-$  > MP  $gurdag$  > NP  $gurda$  ‘kidney’, PIr  $*\acute{u}rka-$  > MP  $gurg$  > NP  $gurg$  ‘wolf’, PIr  $*\acute{u}(i)\acute{j}āna-$  > MP  $gyān$  > 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

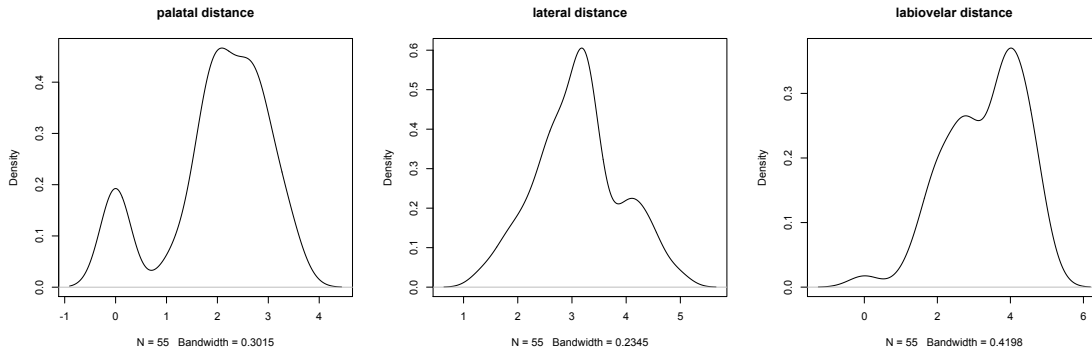
Figure 5.2: Correlation coefficient  $r$  for the relationship between great-circle distance and palatal distance, lateral distance, and labiovelar distance



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

1970). 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



## 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

Table 5.2:  $R^2$  and  $p$  values for correlations between linguistic and geographic distances (asterisks denote significance codes) for Northwest Iranian languages

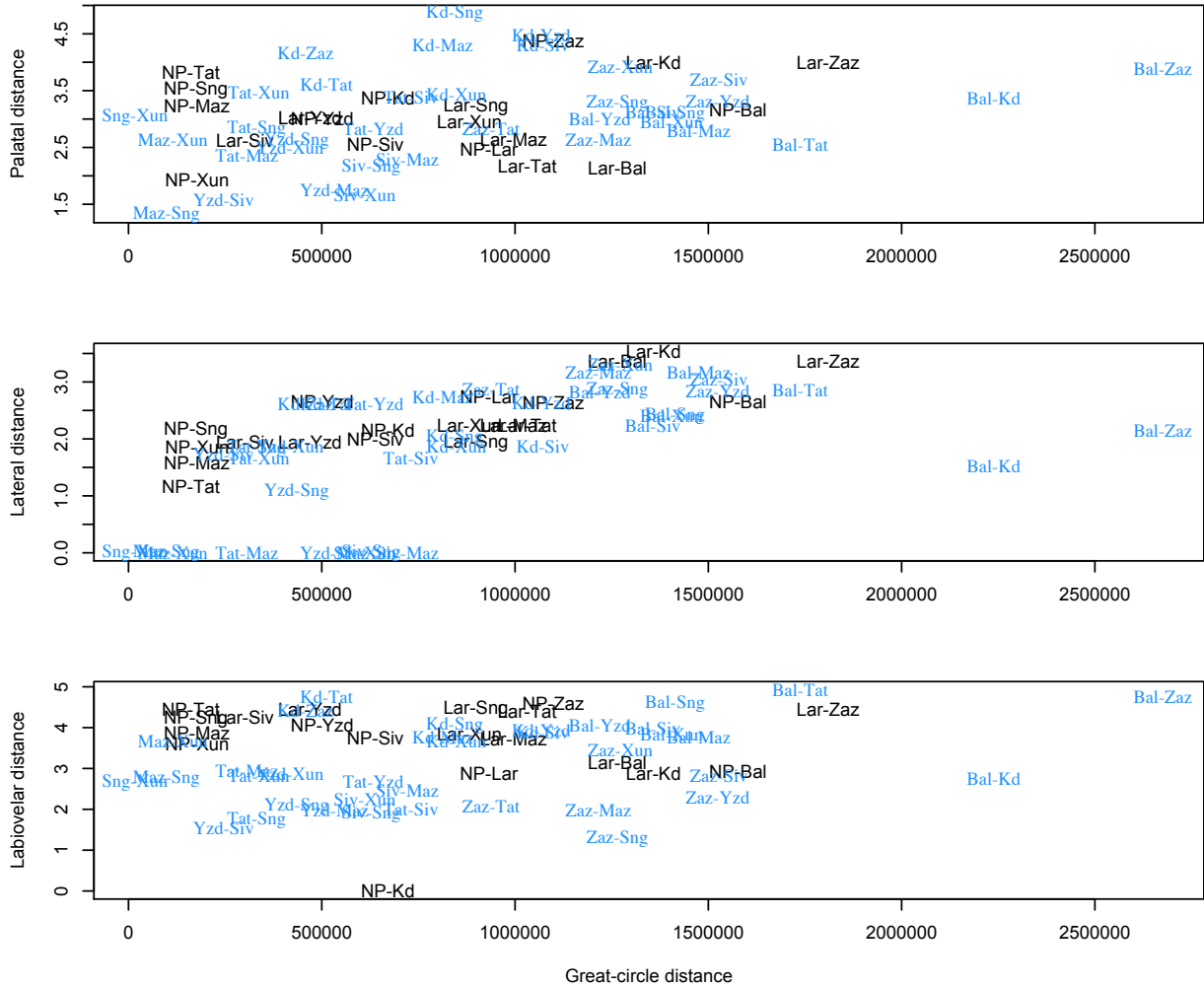
	Great-circle distance		Least-cost distance	
Overall linguistic distance	$R^2 = .26$	$p = .001^{**}$	$R^2 = .19$	$p < .01^{**}$
Palatal distance	$R^2 = .13$	$p = .02^*$	$R^2 = .05$	$p = .18$
Lateral distance	$R^2 = .30$	$p < .001^{***}$	$R^2 = .20$	$p = .005^{**}$
Labiovelar distance	$R^2 = .13$	$p = .02^*$	$R^2 = .16$	$p = .01^*$

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  $*y-$ . While 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 (PIr  $*u\grave{a}ta-$  ‘wind’ serves as an example etymon):

4. Dataset 1 (“VELAR DISTANCE”):

- (a)  $g-$  IN REFLEX OF  $*u\grave{a}ta-$
- (b)  $v-$  IN REFLEX OF  $*u\grave{a}ta-$

5. Dataset 2 (“LABIAL DISTANCE”):

- (c) *b-* IN REFLEX OF  $*_{\underline{u}}\bar{a}ta-$
- (d) *v-* IN REFLEX OF  $*_{\underline{u}}\bar{a}ta-$

Furthermore, since velar distance could potentially represent an interaction between two chronologically separate developments (cf. the early change of PIr  $*_{\underline{u}}- > g-$  /  $_{\underline{r}}$  seen in MP *gurdag*, etc.), I excluded two feature sets pertaining to etyma beginning with the sequence  $*_{\underline{u}}r-$ . 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

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

All West Iranian	Great-circle distance		Least-cost distance	
Velar distance	$R^2 = .12$	$p < .01^{**}$	$R^2 = .14$	$p = .01^{**}$
Labial distance	$R^2 = -.01$	$p = .69$	$R^2 = -.01$	$p = .89$
Only Northwest Iranian				
Velar distance	$R^2 = .27$	$p = .001^{**}$	$R^2 = .31$	$p < .001^{***}$
Labial distance	$R^2 = .02$	$p = .16$	$R^2 = .01$	$p = .22$

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.

## 5.6 Discussion

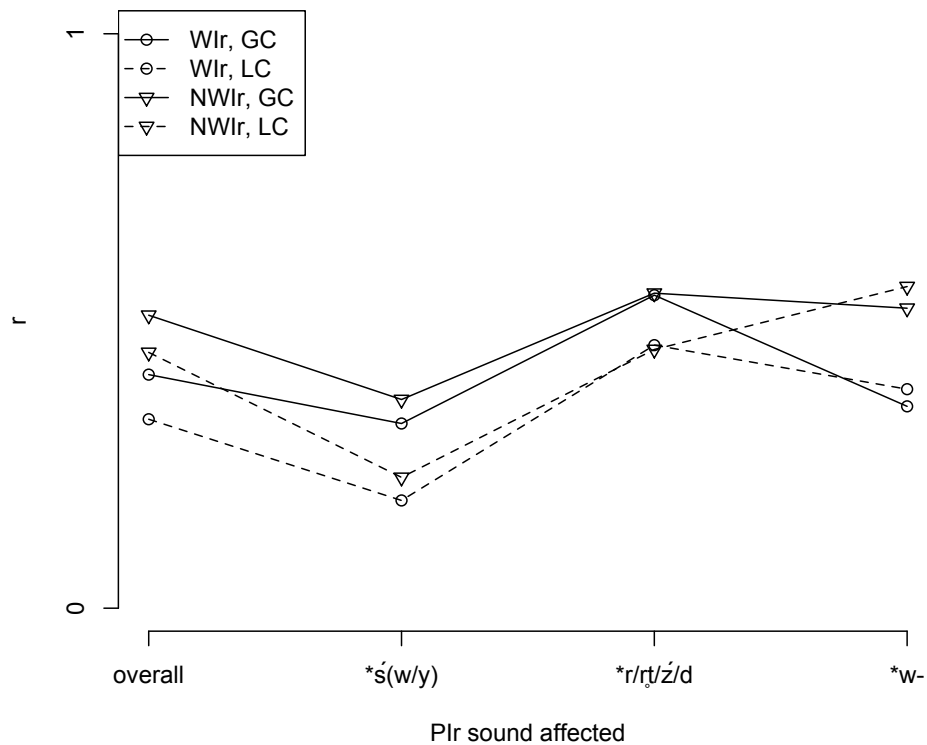
We see the following relationships between the previous section’s results and the chronologies of the actuation and diffusion of these sound changes:

- 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  $*_{\underline{u}}- > g-$  is relatively recent (except before  $*_{\underline{r}}-$ ); 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  $*\underset{\sim}{u}- > b-$  cannot be dated with respect to  $*\underset{\sim}{u}- > g-$  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

Figure 5.5: Correlation coefficient  $r$  for the relationship between geographic distance and overall linguistic distance, palatal distance, lateral distance, and velar distance (excluding reflexes of  $*\underset{\sim}{u}r-$ )



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  $*\underset{\cdot}{u}- > b-$ , 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  $*\underset{\cdot}{u}-$ . The change  $*\underset{\cdot}{u}- > b-$  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  $*\underset{\cdot}{u}- > b$  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  $*\underset{\cdot}{u}-$  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  $< *\underset{\cdot}{u}arna-ka-$ ) show initial *v-*, but appear also to have undergone a sound change known to Persian, OP  $-rn-$   $>$  MP  $-rr-$  (e.g.,  $*\underset{\cdot}{h}uarnah-/ *farnah-$   $>$  MP *xwarrah*  $>$  NP *farr* ‘glory’; cf. also Shirazi *vol* ‘spider’, likely from  $*\underset{\cdot}{u}arna-$ , Schwartz 1971:292, fn. 14). Similarly, *varf* ‘snow’ (PIr  $*\underset{\cdot}{u}afra-$ ), found in Sivandi, Qohrudi, Mazandarani, and Sangesari (where it exists in a doublet with *vafri*), shows the same conservatism in the initial vowel, but the same metathesis ( $-fr > -rf$ ) that is found in NP *barf* ‘*id.*’ ( $< wafri$ ). It is possible that these sound changes ( $-rn- > -rr-$ , 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  $*\underset{\cdot}{u} > b$  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 (Borjani 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.

**I:** *Pashto* ZE (A), (A); *Persian* man (A), (A); *Ormuri* az (A), (A); *Avestan* azəm, məm (A), (A); *Shughni* wuz (A), (A); *Bactrian* 'z (A), (A); *Yaghnobi* man (A), (A); *Wakhi* WUZ (A), (A); *Old Persian* mām, 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);

**all:** *Wakhi* KUXT, CU, CUST (E), (E); *Sogdian* 'γ'c (C), (C); *Balochi* KULL, DRUH, SARO, THEGH, THEWAGH (F), (F); *Khotanese Saka* hama-,harbišša- (A), (A); *Kurdish* giş, hemû, tev (B), (B); *Yaghnobi* háma,yákay (H), (H); *Old Persian* visā (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,pēro (B), (B);

**and:** *Zazaki* ū,kī (B), (B); *Persian* VA (B), (B); *Khotanese Saka* va (B), (B); *Larestani* -o, va (B), (B); *Sogdian* rty (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* -čā, utā (F), (F); *Zebaki* ī,wō (B), (B); *Shughni* atā (F), (F); *Ossetic* aemae (C), (C); *Ishkashmi* za (D), (D); *Parachi* u (B), (B); *Balochi* GUDA, DI, WA, O (B), (B); *Yaghnobi* -at,ham (E), (E);

**animal:** *Balochi* JANWAR, ZANWAR, SHANWAR (H), (A); *Zazaki* heywān (C), (B); *Sogdian* stwrpδyy (I), (I); *Waziri* DZANAWAR (E), (A); *Ossetic* caeraegoj, XAJUAN (G), (G); *Persian* janvar, HEYVAN (A), (A); *Larestani* junovar (D), (A); *Pashto* HAJVAN (F), (B); *Kumzari* haiwan (B), (B); *Shughni* aywūn (J), (B);

- ashes:** *Wakhi* PERG (B), (B); *Pashto* IRA (A), (A); *Sogdian* ’š’kw (A), (A); *Persian* KHAKESTAR (D), (D); *Khotanese Saka* āhāra- (A), (A); *Avestan* ātriiōm (A), (A); *Balochi* PHUR (B), (B); *Waziri* TRA (A), (A); *Kurdish* xwelî (H), (G); *Yaghnobi* xoxistār (G), (D); *Bandari* pūr (B), (B); *Parachi* bhâγ (C), (C); *Zazaki* wel (H), (G); *Middle Persian* dwrystr (I), (F); *Larestani* bal-e garma (J), (E); *Ormuri* xâkistār,yānak (E), (D); *Ossetic* FAENYK (F), (F); *Ishkashmi* usur (A), (A); *Shughni* θīr (A), (A);
- 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* küla (E), (E); *Middle Persian* pwSt (A), (A); *Larestani* kamar, gorda (D), (D); *Sogdian* prc(h) (A), (A); *Parthian* pwSt (A), (A); *Zebaki* dam,med (B), (B); *Ishkashmi* kamuk (G), (G); *Waziri* SHAMZAI (F), (F); *Avestan* parštō (A), (A); *Ormuri* pēc,(pūšt) (A), (A); *Kumzari* kāmār (C), (C); *Shughni* dām (B), (B); *Ossetic* faesontae, C"YLDYM (A), (A); *Pashto* SA (F), (F); *Khotanese Saka* palšti- (A), (A); *Zazaki* pāštī (A), (A);
- bad:** *Avestan* akō, 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* gandā,šakki (B), (B); *Yaghnobi* ganda,bad,šum (B), (B); *Old Persian* gasta (B), (B); *Ossetic* aevzaer (H), (F); *Kumzari* ban’jeh (D), (E); *Ishkashmi* šak (C), (G); *Sogdian* βyz- (J), (H); *Middle Persian* wd (D), (I); *Parachi* bad (I), (E); *Ormuri* γanj,xarâb (B), (B); *Waziri* KHAROP, BAD (F), (D);
- bark:** *Waziri* PATIKAI (D), (D); *Larestani* püss-e derax (A), (A); *Pashto* POST (A), (A); *Shughni* čilyak, pust, anōq (F), (F); *Kurdish* qal, qaşik (C), (C); *Wakhi* DERUXTE PIST (A), (A); *Persian* PUST (A), (A); *Sogdian* c’nwt (E), (E); *Ossetic* c’ar (E), (E); *Yaghnobi* pust,pustloq (A), (A); *Balochi* GAWAZ (B), (B);
- because:** *Shughni* čarō didi (B), (B); *Pashto* DZEKA CE, VALI CE (B), (B); *Persian* CHUN (B), (B); *Balochi* PHA HAW-AN SAUAV, PHA HAW-AN KHAN, KI (C), (C); *Zazaki* welo-k (D), (D); *Ossetic* umaen aemae (E), (E); *Wakhi* CIZER (B), (B); *Zebaki* ke, tsīz-bā ke (B), (B); *Avestan* zī (A), (A);
- belly:** *Balochi* LAF (C), (C); *Waziri* GADOLYAI (G), (G); *Avestan* udarəm (A), (A); *Ishkashmi* dēr (A), (A); *Kurdish* zik (D), (D); *Sogdian* kδ’r’k (A), (A); *Middle Persian* prwdg? (A), (A); *Kumzari* shukum (F), (F); *Zebaki* dēr (A), (A); *Yaghnobi* dāra (A), (A); *Wakhi* WANJ, WERD, DOR (B), (B); *Larestani* ‘aškam (F), (F); *Khotanese Saka* aha,ūra,garba- (A), (A); *Parachi* aštaf,xīṭ (I), (I); *Shughni* qīč (J), (J); *Ossetic* guybyn (H), (H); *Zazaki* pīze (E), (E); *Persian* SHEKAM (F), (F); *Pashto* GEDA, NAS, XETA (G), (G); *Ormuri* škamba (stomach) (F), (F);



**big:** *Zebaki* kata (H), (H); *Shughni* γullā,xidīr (H), (H); *Larestani* gap (E), (E); *Wakhi* LUP (B), (B); *Balochi* MAZAAN, MAZ EN (A), (A); *Middle Persian* wzrg (D), (D); *Yaghnobi* kátta (H), (H); *Zazaki* gird,pīl,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, çivik (D), (D); *Waziri* MARGHAI (C), (C); *Zazaki* miričik,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); *Kumzari* 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* mrγ’ (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 GEREF TAN (B), (B); *Wakhi* DENDUK DI-, GUP DI- (B), (B); *Sogdian* zβ’t- (I), (I); *Kumzari* kha’adish (pst) (E), (E); *Shughni* pirēnd (L), (L); *Avestan* daš- (A), (A); *Parthian* gšt (J), (J); *Middle Persian* gc (J), (J); *Pashto* CICEL (G), (G); *Balochi* WARAGH, WARTHA (C), (C); *Waziri* CHICHEL (G), (G); *Kurdish* geztin (D), (D); *Larestani* kap kanda, gereta (F), (F); *Ossetic* xaecyn (H), (H); *Khwarezmian* bγ’h- (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* š’w (A), (A); *Kurdish* reş (B), (B); *Avestan* siiāuua-, 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* īn,šun (A), (A); *Wakhi* XUN, WUSEN (A), (A); *Kumzari* khūwaym (A), (A); *Parthian* gwxn (A), (A); *Avestan* vohuni (A), (A); *Zazaki* gūnī (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* vāiti (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* haḡḡ (B), (B); *Ossetic* staeg (A), (A); *Kumzari* khār (C), (C); *Sogdian* 'stk- (A), (A); *Ormuri* stoγān (A), (A); *Pashto* HED, HADUKAJ (B), (B); *Avestan* ast-, as (A), (A); *Larestani* 'ossoxun (A), (A); *Shughni* sitxun (A), (A); *Zazaki* este (A), (A); *Wakhi* YUSC, USTUXON (A), (A); *Yagnobi* sitak (A), (A); *Balochi* HAD (B), (B); *Waziri* HADIKAI (B), (B); *Persian* ostokhan (A), (A);
- breast:** *Yagnobi* vūna,cic (F), (F); *Zazaki* sēnē,virār (C), (C); *Khwazmian* (y')ft'n (A), (A); *Kurdish* pēsîr, çiçik (B), (B); *Persian* pistān (B), (B); *Zebaki* bar (E), (E); *Sogdian* 'stnh (A), (A); *Khotanese Saka* tcījsa (G), (G); *Kumzari* sīnō (C), (C); *Shughni* bar, bat, sīnā, biš (E), (E); *Ishkashmi* pēšbar,cīcī (E), (E); *Parachi* sīnā (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 (BREATH) (B), (B); *Persian* NAFAS KASHIDAN (D), (D); *Ossetic* ULAEFYN (G), (G); *Shughni* xu dām ziwestow (H), (H); *Khotanese Saka* dam- (H), (H); *Larestani* nafas kerda (E), (E); *Zazaki* nefes (breath) (C), (C);
- burn:** *Avestan* dažiiete (A), (A); *Pashto* SVADZEDEL, SVADZEL (D), (D); *Zazaki* vešāyiš (E), (E); *Waziri* BALEDEL, SWEL (C), (C); *Larestani* sota (D), (D); *Khwazmian* 'βr'z- (F), (F); *Middle Persian* h'w (E), (E); *Yagnobi* suč- (D), (D); *Khotanese Saka* suv-,dajs- (A), (A); *Persian* SUKHTAN (D), (D); *Ossetic* sudzyn (D), (D); *Parthian* h'w,bry- (E), (E); *Shughni* sūzaḡ čīdōw (D), (D); *Kurdish* sotîn, şewtî (D), (D); *Parachi* thēw- (B), (B); *Wakhi* THAU- (B), (B); *Balochi* BALAGH (C), (C); *Sogdian* swc (D), (D);
- child:** *Persian* BACHCHE (D), (D); *Parthian* fryznd (M), (J); *Avestan* apərənāiiukō (A), (A); *Zazaki* qeçek,gede,leyr,domān,qiž,tūt (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* bačā,kūdak,tifl (B), (B); *Parachi* bačī (I), (D); *Yagnobi* pulla,bača,gudak,farzand (K), (D); *Sogdian* 'jwn (L), (J); *Wakhi* KUDUK, ZA, ZUMAN (B), (B); *Waziri* WORKAI (F), (F); *Middle Persian* fryznd (M), (J); *Balochi* CHUKH (C), (C); *Ishkashmi* zāman (J), (I);
- cloud:** *Ishkashmi* gulbāduk (H), (H); *Larestani* 'abr (A), (A); *Parthian* myg (G), (G); *Shughni* abri (A), (A); *Pashto* TORA URIADZ, XERA URIADZ (F), (F); *Ossetic* mig' (G), (G); *Kurdish* ewr (A), (A); *Yagnobi* abr,tira (A), (A); *Waziri* WERYEZ (E), (E); *Avestan* aβrəm, snaoδō, maēγəm (A), (A); *Kumzari* nīm (< Ar?) (D), (D); *Yazgulyam* varm (I), (I); *Sogdian* pr'yβ'k (A), (A); *Ormuri* yēwər (A), (A); *Parachi* a'ir (A), (A); *Wakhi* MOR, WETIS (B), (B); *Persian* ABR (A), (A); *Khotanese Saka* pryaura (A), (A); *Balochi* JH UR (C), (C);
- cold:** *Pashto* SOR (A), (A); *Avestan* sarətō, aotō (A), (A); *Balochi* SARTH, GWAHAR (A), (A); *Ossetic* uazal (B), (B); *Kurdish* sar (A), (A); *Parachi* eštāwo (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* cāk (C), (C); *Yaghnobi* sort,xunuk (A), (A); *Bashkardi* sūrt (A), (A); *Waziri* SOR (A), (A); *Shughni* sard (A), (A); *Zazaki* serd (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 aēiti (A), (A); *Kumzari* hāmed (pst) (C), (C); *Zazaki* āmiyāyiš (C), (C); *Middle Persian* 'md- (C), (C); *Wakhi* WEZI- (A), (A); *Yaghnobi* vvow- (G), (G); *Bactrian* aga- (A), (A); *Sogdian* "γt- (A), (A); *Zebaki* is- (F), (F); *Kurdish* hatin (B), (B); *Old Persian* ā-jam-, āitij (A), (A); *Parthian* 'gd (A), (A); *Persian* AMADAN (C), (C); *Ishkashmi* āGad (A), (A); *Pashto* RATLEL (D), (D);
- count:** *Khotanese Saka* Sumār (A), (A); *Persian* SHOMORDAN (A), (A); *Shughni* asōb (B), (B); *Ossetic* nymajyn (A), (A); *Sogdian* ptšmr (A), (A); *Zazaki* āmordiš (A), (A); *Pashto* SMEREL (A), (A); *Parthian* 'šm'r ('number') (A), (A); *Balochi* GANNAGH, GANNITHA (C), (C); *Waziri* GANREL (C), (C); *Middle Persian* 'šm'r- (A), (A); *Wakhi* ISOB TSER- (B), (B);
- cut:** *Persian* BORIDAN (E), (E); *Avestan* kərəntaiti (A), (A); *Old Persian* \*kart- (A), (A); *Balochi* CHAKAGH, CHAKITHA (C), (C); *Parachi* mač (F), (F); *Khotanese Saka* ttäS- (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* tēb (J), (J); *Larestani* boleda (E), (E); *Wakhi* RESED- (B), (B); *Parthian* q'f? (I), (I); *Zazaki* qesnāyiš (D), (D); *Waziri* PREKREL (A), (A); *Zebaki* keT- (past ppl) (A), (A); *Kumzari* qašai'kin (imp sg) ( ), ( );
- day:** *Sogdian* myθ (D), (D); *Avestan* aiiarə, 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); *Parachi* dewās,ruč (B), (B); *Middle Persian* rwc,rwz (B), (B); *Ormuri* rōž (B), (B); *Waziri* VREZ, WREZ (B), (B); *Pashto* VRADZ (B), (B); *Yazgulyam* miθ (D), (D); *Old Persian* rauča (B), (B); *Shughni* mēθ, rüz (D), (D); *Wakhi* ROR, REWOR (B), (B); *Larestani* rüz roz (B), (B); *Zebaki* mī (D), (D); *Persian* RUZ (B), (B); *Bandari* rüz (B), (B); *Khwarezmian* myθ (D), (D);
- die:** *Middle Persian* myr- (A), (A); *Avestan* miriiete (A), (A); *Wakhi* MERI- (A), (A); *Sogdian* myr (A), (A); *Zebaki* murum (A), (A); *Zazaki* merdiš (A), (A); *Shughni* Mar (A), (A); *Yaghnobi* mir-, marg vu- (A), (A); *Old Persian* mariyataiy (A), (A); *Parachi* mer- (A), (A); *Persian* MORDAN (A), (A); *Parthian* myr- (A), (A); *Kumzari* murd (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* maelyn (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* kūč, kuht (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-ék (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* čān (A), (A); *Pashto* KINEL, KINDEL (A), (A); *Waziri* KANDEL (A), (A); *Parachi* kusēw (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* čālīn, čirkin, γažd, qarq (I), (I); *Sogdian* ”γwstk (L), (L); *Larestani* gana (E), (E); *Ormuri* čirk (I), (I); *Yaghnobi* xira, loynók (K), (K); *Waziri* KHACHEN, KHIRAN (G), (G); *Zazaki* ?imin (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* šve (< Ind?) (C)<sub>2</sub> (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* nuazyn, 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* qōq,xušk (B), (B); *Wakhi* WESK, XUSK (A), (A); *Ishkashmi* kāk (B), (B); *Waziri* WUCH, SIR (A), (A); *Parachi* öškār,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* wōka (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,γubōr,sit (M), (B); *Wakhi* GERD, XSUREM (B), (B); *Parachi* čār (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* paąsnuš (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); *Shughni* γu~γ (A), (A); *Persian* gush (A), (A); *Yaghnobi* Guš (A), (A); *Old Persian* gauša (A), (A); *Bandari* güš (A), (A); *Khotanese Saka* gū (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); *Pashto* GVAZ (A), (A); *Kurdish* guh (A), (A); *Ormuri* gōi (A), (A); *Larestani* goš (A), (A); *Parachi* gū (A), (A); *Middle Persian* gwš, (A), (A); *Wakhi* YIS (A), (A);
- earth:** *Kurdish* xak, erd (E), (E); *Parachi* khex (K), (J); *Balochi* MITTI (D), (D); *Ormuri* xâk (J), (H); *Waziri* KHOVRA, WATAN (G), (G); *Kumzari* zamiyō (def) (A), (A); *Yaghnobi* zoy,zamin (A), (A); *Shughni* sit, zamin, zimad (N), (C); *Larestani* bal (F), (F); *Old Persian* būmiš (B), (B); *Pashto* XAK, MDZEKA (H), (H); *Sogdian* γwrm(h) (M), (I); *Wakhi* XOK, SET (C), (C); *Persian* ???? (A), (A); *Avestan* zâ (A), (A); *Zazaki* heṛ (E), (E); *Ossetic* zaexx, SYDZYT (I), (A); *Khotanese Saka* uysmä,ššandā (C), (C); *Ishkashmi* šit (< Wakhi?) (L), (C);
- eat:** *Kurdish* xwarin (A), (A); *Pashto* XVAREL (A), (A); *Wakhi* YAU- (B), (B); *Zazaki* werdiš (A), (A); *Persian* KHORDAN (A), (A); *Avestan* xvaraiti (A), (A); *Ormuri* xr-, xr-; *Waziri* KHWAREL (A), (A); *Kumzari* khōr (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* hêk (B), (B); *Ossetic* ajk (B), (B); *Pashto* HAGEJ (B), (B); *Wakhi* TUXM MURGH (A), (A); *Ormuri* supāl (D), (D); *Kumzari* khaig (B), (B); *Persian* tokhm (A), (A); *Parachi* ēx (B), (B); *Khwarezmian* y’k (B), (B); *Balochi* HAIKH, ANU (B), (B); *Waziri* YOWYA, YIYA (B), (B); *Shughni* tarmurx (A), (A); *Khotanese Saka* āhā (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); *Shughni* čaxm, cem, dīda (A), (A); *Bandari* čehem (A), (A); *Khotanese Saka* tceiman- (A), (A); *Zazaki* čim (A), (A); *Middle Persian* cšm (A), (A); *Wakhi* CEZM (A), (A); *Ishkashmi* tsām (A), (A); *Balochi* CHHAM (A), (A); *Persian* chasm (A), (A); *Zebaki* tsām (A), (A); *Yaghnobi* Gurda,wenna (D), (D); *Ossetic* caest (A), (A); *Old Persian* čašna (A), (A); *Bashkardi* čehm (A), (A); *Larestani* caš (A), (A); *Ormuri* cimī (A), (A); *Waziri* STERGA (B), (B); *Kurdish* çav (A), (A); *Parthian* cšm (A), (A); *Avestan* čama (A), (A); *Parachi* dīda (C), (C); *Sogdian* cšm (A), (A);
- fall:** *Zazaki* kewtiš,gunāyiš (C), (C); *Yaghnobi* dewi-,tiraš-,ruš- (H), (H); *Zebaki* idāwī (H), (H); *Parthian* qp- (C), (C); *Khotanese Saka* kas- (C), (C); *Persian* OFTADAN (D), (D); *Avestan* pataiti (A), (A); *Ormuri* γūz (G), (G); *Kumzari* keft (pst) (C), (C); *Balochi* KHAFAGH, KHAPTA (C), (C); *Sogdian* w’pt (A), (A); *Wakhi* PERVE-, PULUN-, WUZ- (B), (B); *Shughni* raz (I), (I); *Pashto* LVEDEL (F), (F); *Waziri* 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* dūr (A), (A); *Shughni* dar, dūr (A), (A); *Parthian* dwr (A), (A); *Ossetic* dard (B), (B); *Wakhi* DIR (A), (A); *Balochi* DIR (A), (A); *Avestan* dūraē (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); *Zebaki* dīr (A), (A); *Parachi* durīn (A), (A); *Zazaki* dūrī (A), (A);

**fat:** *Pashto* GVARI (F), (F); *Avestan* pīuuō, āzūitiš (A), (A); *Balochi* PHIGH (B), (B); *Parachi* čarbū (E), (E); *Ossetic* soj (G), (G); *Yaghnobi* ruGen (D), (D); *Persian* charbi (E), (E); *Larestani* carbi (E), (E); *Khotanese Saka* päyā,tcārba- (E), (E); *Kurdish* qelew (C), (C); *Shughni* čarvi, zōγ (E), (E); *Waziri* WOZDA (A), (A); *Sogdian* crp,rwγn (E), (E); *Zazaki* ruwen (D), (D);

**father:** *Waziri* PLOR, BABA (A), (A); *Zebaki* 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); *Shughni* dōd, padar, pid (A), (A); *Balochi* PHITH, PITH (A), (A); *Sogdian* (')ptr (A), (A); *Ishkashmi* tot,tāt (B), (B); *Zazaki* pī,bābī (A), (A); *Kumzari* bap (C), (C); *Avestan* pita (A), (A); *Larestani* buvā, babā (C), (C); *Parthian* pyd (A), (A); *Wakhi* TUT (B), (B); *Bandari* bap (C), (C); *Old Persian* pitā (A), (A); *Pashto* PLAR (A), (A);

**fear:** *Shughni* aks čīdōw (F), (F); *Parachi* bīm (n.) (A), (A); *Yaghnobi* čukáyr (D), (D); *Pashto* BEREDDEL (A), (A); *Old Persian* tarsatiy (B), (B); *Waziri* DAREDEL, WYEREDDEL (B), (B); *Avestan* biβāiia (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* tursidish (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* parənəm (A), (A); *Zazaki* pūrtī (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* BENA, BANEKA (C), (C);

**few:** *Pashto* LEZ (B), (B); *Avestan* kamna- (A), (A); *Yaghnobi* kam,barg,andak (A), (A); *Shughni* andák (E), (E); *Zazaki* tāy(n),kemī,senik (A), (A); *Zebaki* tsamend (D), (D); *Sogdian* kβn- (A), (A); *Wakhi* KUM (A), (A); *Balochi* KHARDE, KHAM (A), (A); *Ossetic* cysyl (C), (C); *Larestani* kam (A), (A); *Old Persian* kamnā (A), (A); *Persian* KAM (A), (A); *Waziri* LEZH, LEZHKI (B), (B);

**fight:** *Pashto* DZANGEDEL (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* qastīn anjīvdōw (I), (I);

*Avestan* pərətaitē, yūdiieiti (A), (A); *Waziri* JANG, JAGGARRA BALWA (C), (C); *Sogdian* ”x’s (H), (H); *Wakhi* JUNG TSER- (C), (C);

**fingernail:** *Shughni* nōxūn (B), (B); *Khotanese Saka* nāhun- (B), (B); *Kurdish* neynūk (B), (B); *Ishkashmi* ingituk (D), (D); *Bandari* penj (C), (C); *Avestan* srauuō (pl.) (A), (A); *Khwarezmian* šwk (A), (A); *Kumzari* nikhin (B), (B); *Sogdian* n’γ’n (B), (B); *Larestani* nāxon (B), (B); *Ormuri* taxt ta anguš (D), (D); *Yaghnobi* naxna (B), (B); *Zazaki* nengū (B), (B); *Persian* N?XUN (B), (B);

**fire:** *Khotanese Saka* dai (C), (C); *Kumzari* hātish (A), (A); *Sogdian* ”tr (A), (A); *Parachi* â? (A), (A); *Shughni* alōw, yōc (A), (A); *Ishkashmi* rōšni (B), (B); *Avestan* ātarš (A), (A); *Persian* ATASH (A), (A); *Zebaki* rošmī (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* \*āci- (A), (A); *Parthian* ’dwr (A), (A); *Zazaki* ādir (A), (A); *Yaghnobi* olow, (A), (A); *Yazgulyam* yets (A), (A); *Wakhi* RUXUNIGH (B), (B); *Pashto* OR (A), (A); *Kurdish* agir (A), (A); *Middle Persian* ’dwr (A), (A);

**fish:** *Shughni* mōyi (loan) (E), (A); *Yaghnobi* mahi (loan) (F), (A); *Balochi* MAHI (G), (A); *Larestani* ma’i (A), (A); *Wakhi* KUP (B), (B); *Avestan* masiō (A), (A); *Ossetic* kaesag (C), (C); *Khotanese Saka* kava (B), (B); *Parthian* m’sy’g (A), (A); *Kurdish* masī (A), (A); *Pashto* KAB (B), (B); *Parachi* masō (A), (A); *Zazaki* māse (A), (A); *Persian* mahi (A), (A); *Middle Persian* m’hyg (A), (A); *Kumzari* mī (A), (A); *Bandari* müyi (A), (A); *Sogdian* kp’ (B), (B); *Ormuri* (mā’i) (D), (A);

**five:** *Avestan* panča (A), (A); *Ossetic* fondz (A), (A); *Khotanese Saka* paMjsa- (A), (A); *Zebaki* pūnz (A), (A); *Wakhi* PANZ (A), (A); *Yaghnobi* panč,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* pūnz (A), (A); *Parthian* pnj (A), (A); *Shughni* panj,pinž (A), (A); *Parachi* pönž (A), (A); *Larestani* panj (A), (A); *Yazgulyam* pindz (A), (A); *Middle Persian* pnz,pnc (A), (A);

**float:** *Waziri* BAIYEDEL (C), (C); *Larestani* ‘a lü-’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* y’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* rižiyāyīš (A), (A); *Pashto* BAHEDDEL, TOJEDEL (D), (D); *Khotanese Saka* ttajs- (G), (G); *Avestan* γžāraiti, raodaiti (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,čičēge (C), (A); *Shughni* gul (O), (A); though gul < \*wRda) (D), (A); *Middle Persian* ‘sprhm, (A), (D); *Parachi* gul (K), (A);

- Sogdian* 'sp'rγmy(y) (A), (D); *Larestani* gol (E), (A); *Wakhi* GUL, SPREGH (A), (D); *Bandari* gol (F), (A); *Persian* GOL (D), (A); *Kumzari* ward (< Ar?, Ar?; *Yaghnobi* gul (M), (A); *Khotanese Saka* spātaa- (N), (E); *Ormuri* gul (J), (A); *Parthian* w'r (D), (A);
- fly:** *Parachi* rhâz- (D), (D); *Avestan* frauuaite (pataiti Daevic) (A), (A); *Waziri* WRATEL (A), (A); *Sogdian* frwz- (D), (D); *Persian* PARVAZ KARDAN (D), (D); *Pashto* ALVOTEL (E), (E); *Shughni* riwâz (D), (D); *Balochi* BAL GIRAGH, BAL GIPTA (B), (B); *Middle Persian* prwz- (D), (D); *Zazaki* peṛāyiš, fir-dāyiš (C), (C); *Parthian* frwz- (D), (D); *Larestani* parda, fal kerda (D), (D); *Kurdish* firrîn (C), (C); *Ossetic* taexyn (F), (F); *Yaghnobi* par-, furr- (D), (D);
- fog:** *Ossetic* mig' (C), (C); *Yaghnobi* tuman (G), (G); *Waziri* BADAL, LERA (F), (F); *Persian* MEH (C), (C); *Parthian* nyzm'n (D), (D); *Pashto* MIH (C), (C); *Avestan* dunma (A), (A); *Balochi* DITHLO (B), (B); *Bandari* bāpir (E), (E); *Larestani* nezvā (D), (D); *Shughni* manyōr, tafaḡ, tuman (G), (G);
- foot:** *Old Persian* pāda (A), (A); *Bandari* kap-e pā (A), (A); *Kumzari* pā (A), (A); *Waziri* PSHA (A), (A); *Zebaki* pūd (A), (A); *Ormuri* pāi (A), (A); *Avestan* pāda (du.) (A), (A); *Balochi* PHADH (A), (A); *Khotanese Saka* pā (A), (A); *Parthian* p'd (A), (A); *Persian* pa (A), (A); *Middle Persian* p'y (A), (A); *Wakhi* PUED (A), (A); *Ishkashmi* pu (A), (A); *Pashto* PSA (A), (A); *Khwarezmian* p'δ (A), (A); *Bashkardi* pū (A), (A); *Sogdian* p'δ (A), (A); *Zazaki* pā (A), (A); *Kurdish* pê (A), (A); *Ossetic* fad, K"AX (B), (B);
- four:** *Wakhi* TSEBUR, SUBUR (A), (A); *Balochi* CHIAR (A), (A); *Yazgulyam* cer (A), (A); *Ishkashmi* tsafur (A), (A); *Parthian* cf'r (A), (A); *Ossetic* syppar (A), (A); *Zebaki* tsafur (A), (A); *Yaghnobi* tifer, čor (A), (A); *Larestani* cār (A), (A); *Ormuri* cār (A), (A); *Middle Persian* ch'r (A), (A); *Pashto* CALOR (A), (A); *Khwarezmian* cf'r (A), (A); *Sogdian* ctf'r (A), (A); *Waziri* TSALOR, TSALWOR, TSALWER (A), (A); *Shughni* čor, cavōr (A), (A); *Avestan* čaθbārō (A), (A); *Kumzari* chār (A), (A); *Khotanese Saka* tcāMrai (A), (A); *Persian* CHAHAR (CHAR) (A), (A);
- freeze:** *Wakhi* YIS WOTS-, YIS TSER- (A), (A); *Persian* YAKH BASTAN (A), (A); *Balochi* MADHAGH, MASTAGH (B), (B); *Pashto* JAX KEDEL (A), (A); *Ossetic* IX KAENYN, saelyn (A), (A); *Waziri* KARANG (FROZEN) (C), (C);
- fruit:** *Khotanese Saka* varga- (J), (D); *Middle Persian* b'r, myw (J), (D); *Parachi* mēwa (I), (A); *Sogdian* βryy (J), (D); *Wakhi* MIWA (A), (A); *Khwarezmian* βrk (J), (D); *Bactrian* βrg (J), (D); *Ossetic* dyrg' (G), (B); *Balochi* MEWA, BAR (B), (A); *Persian* mive (C), (A); *Ormuri* mēwa (fruit) (H), (A); *Larestani* miva (D), (A); *Waziri* MEWA (E), (A); *Pashto* MEVA (F), (A); *Yaghnobi* pula (J), (C);
- full:** *Sogdian* pwrn (A), (A); *Shughni* čōq, pur, purrā (A), (A); *Avestan* pərənō (A), (A); *Kurdish* pirr, tijî (A), (A); *Parthian* pwr (A), (A); *Middle Persian* pwr (A), (A); *Zazaki* piṛ (A), (A); *Parachi* tha? (B), (B);



- give:** *Kurdish* dan (A), (A); *Shughni* dedow (A), (A); *Balochi* DEAGH, DATHA (A), (A); *Avestan* daḏāiti (A), (A); *Pashto* VERKAVEL (B), (B); *Ossetic* daettyn (A), (A); *Persian* DADAN (A), (A); *Khwarezmian* hβr (E), (E); *Zazaki* dāyiš (A), (A); *Parachi* baxš kan- (C), (C); *Bandari* dāden (A), (A); *Yaghnobi* tifār- (D), (D); *Ishkashmi* da- (A), (A); *Middle Persian* dy- (A), (A); *Waziri* DERKREL (A), (A); *Bactrian* lh- (A), (A); *Zebaki* da- (A), (A); *Old Persian* dadātiy (A), (A);
- good:** *Sogdian* šr,xwp (C), (C); *Bandari* xub (B), (B); *Yaghnobi* obadon (J), (J); *Old Persian* \*vahu-, naiba (A), (A); *Balochi* šarr (C), (C); *Zazaki* weš,hol,rind (D), (D); *Pashto* SA (C), (C); *Ishkashmi* fri,(nēk) (I), (I); *Parachi* kārī,nēk,xub (H), (H); *Ossetic* xorz, DZAEBAEX (G), (G); *Khwarezmian* xwb (B), (B); *Shughni* bašānd (D), (D); *Khotanese Saka* śsāra (C), (C); *Wakhi* XUB, BUF (B), (B); *Zebaki* ferī (I), (I); *Larestani* xuš (F), (F); *Middle Persian* xwb (B), (B); *Avestan* va?huš (A), (A); *Kumzari* khair (< Ar) (E), (E); *Persian* KHUB (B), (B); *Ormuri* širr (C), (C); *Kurdish* baş, qenc, rind (D), (D); *Waziri* SHE (C), (C);
- grass:** *Parachi* alafī,gihāi,sauza (H), (C); *Waziri* WOSHE (A), (A); *Persian* ALAF (C), (C); *Larestani* ‘alaf (F), (E); *Shughni* alaf, wōx (A), (A); *Khotanese Saka* ggīsai (D), (D); *Pashto* VASE (A), (A); *Kumzari* gīyā (D), (D); *Wakhi* WUS (A), (A); *Bashkardi* gādā(h), gāda (< Balochi) (E), (D); *Ormuri* γwāšī (A), (A); *Middle Persian* gy’w,mrw (D), (D); *Zazaki* vāš (A), (A); *Ishkashmi* ūš (A), (A); *Bactrian* wš (A), (A); *Balochi* RENV, REM (B), (B); *Khwarezmian* wš (A), (A); *Yaghnobi* wayš,marG (A), (A); *Ossetic* kaerdaeg (G), (G);
- green:** *Waziri* SHIN, ZARGHIN (C), (B); *Ossetic* c’aex (I), (D); *Kurdish* hēşm, kesk (C), (B); *Bandari* süz (H), (A); *Wakhi* SUVZ (A), (A); *Persian* SABZ (E), (A); *Parthian* hwzrgwn (M), (F); *Zazaki* kiḥo,?ewi (D), (C); *Pashto* SIN (C), (B); *Yaghnobi* Gūra (L), (E); *Parachi* sābz (J), (A); *Middle Persian* hwzrgwn (M), (F); *Larestani* sowz (G), (A); *Shughni* savž, xum (O), (A); *Khotanese Saka* gvā- (N), (G); *Sogdian* ‘xs’yn (C), (B); *Kumzari* sauz (F), (A); *Balochi* SAVZ, MAUNSHAR (B), (A); *Ishkashmi* (sabz) (K), (A);
- guts:** *Balochi* ROTH, RODH (B), (B); *Pashto* KULMA (C), (C); *Persian* RUDE (B), (B); *Waziri* KULMA, LARMIN (C), (C); *Ossetic* t’ang (D), (D); *Wakhi* SINGER (A), (A); *Larestani* rüda (B), (B); *Shughni* darūn (C), (C); *Parachi* rüdā (C), (C);
- hair:** *Zebaki* seGund (H), (H); *Bandari* müd (D), (D); *Balochi* PHUT (C), (C); *Shughni* mūy, tōr, γūnj (D), (D); *Parachi* dōš,jāl (F), (F); *Zazaki* mū,por,pinč (D), (D); *Yaghnobi* diraw,pašm (F), (F); *Larestani* mü (D), (D); *Persian* mu (D), (D); *Pashto* VESTE (A), (A); *Ishkashmi* Gēnuk (G), (G); *Ossetic* SAERYX"UYN, x’uyn (E), (E); *Kurdish* mū, por (D), (D); *Sogdian* wrs,žw (A), (A); *Wakhi* SUFS, PICA (B), (B); *Waziri* WESHTE (A), (A); *Kumzari* mū (D), (D); *Khotanese Saka* drau,-ggūna- (F), (F); *Avestan* varəšō (A), (A);

**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); *Khwazmian* (y')δst (A), (A); *Shughni* dast, dust, xikufta (A), (A); *Avestan* zastō (A), (A); *Old Persian* dasta (A), (A); *Ormuri* dest (A), (A); *Persian* dast (A), (A); *Zazaki* dest (A), (A); *Kumzari* dist (A), (A); *Balochi* DAST (A), (A); *Zebaki* dāst (A), (A); *Kurdish* dest (A), (A); *Khotanese Saka* ggoštā (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* 'ána (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:** *Wakhi* SER (A), (A); *Kumzari* sōr (A), (A); *Parthian* sr (A), (A); *Zazaki* sere (A), (A); *Yaghnobi* kallá,sarkallá,sar (B), (B); *Larestani* sera (A), (A); *Parachi* sōr (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* sōr (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* bihištīn (C), (C); *Middle Persian* ngwš (E), (E); *Pashto* ARVEDEL (C), (C); *Ishkashmi* šud (pst) (A), (A); *Avestan* surunaoti (A), (A); *Parthian* ngwš,'zgwł (E), (E); *Sogdian* ptγwš- (E), (E); *Waziri* WORWEDEL, ARWEDEL (C), (C); *Balochi* ASKHANAGH, ASKHUTHA (B), (B); *Larestani* šonofta, goš kerda (B), (B); *Bandari* ešnūten (B), (B); *Ossetic* x'ycyn (D), (D); *Zazaki* āšnāwitiš (B), (B); *Shughni* xīn (A), (A);

**heart:** *Ormuri* zlī (A), (A); *Kumzari* dil (A), (A); *Khotanese Saka* ysaraka (A), (A); *Pashto* ZRE (A), (A); *Wakhi* PEZUV (B), (B); *Ossetic* zaerdae (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, zeṛī (A), (A); *Sogdian* (γrdy'),δrjyy (A), (A); *Bandari* del (A), (A); *Bashkardi* der (A), (A); *Yaghnobi* dil (E), (A); *Khwazmian* zrz (A), (A); *Avestan* zərəd- (A), (A); *Persian* QALB (C), (C); *Zebaki* āuzak (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* sangī (B), (B); *Waziri* DRIND, SAKHT (C), (C); *Sogdian* β'rcyk (E), (E); *Shughni* wazmin, zūr (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* wazmīn (D), (D);

**here:** *Ossetic* am ( ), (D); *Shughni* arēd ( ), (F); *Waziri* DELE, DOLATA ( ), (C); *Zazaki* tiyā,injā,nāzā ( ), (A); *Balochi* EDHA ( ), (A); *Wakhi* YEM, YAEI, DREM, HADREM

- ( ), (B); *Yaghnobi* idoka ( ), (A); *Pashto* DELTA, DALE ( ), (C); *Sogdian* 'mδ ( ), (D); *Avestan* idā ( ), (A); *Parachi* ēk ( ), (E); *Kumzari* aywo/aiyo ( ), (A); *Larestani* 'inke ( ), (A); *Persian* INJA ( ), (A); *Khotanese Saka* ttattī- ( ), (H); *Zebaki* mādak ( ), (G); *Old Persian* idā ( ), (A);
- hit:** *Ormuri* zan- (A), (A); *Middle Persian* jn- (A), (A); *Balochi* JANAGH (A), (A); *Zazaki* dāyiš...-ro,piro-dāyiš,kuwāyiš (C), (C); *Ossetic* cavyn, NAEMYN (E), (E); *Larestani* zata (A), (A); *Khotanese Saka* jse- (A), (A); *Parachi* jan- (A), (A); *Old Persian* jantiy (A), (A); *Parthian* jn- (A), (A); *Persian* ZADAN (A), (A); *Avestan* jainti (A), (A); *Bashkardi* zan (A), (A); *Sogdian* zn (A), (A); *Pashto* VAHEL (D), (D); *Wakhi* DI- (B), (B); *Waziri* WAHEL, TAKAWEL (D), (D);
- hold:** *Parachi* ši (E), (E); *Persian* GEREFTAN (B), (B); *Sogdian* d'r (A), (A); *Avestan* dāraieiti (A), (A); *Shughni* anjav (F), (F); *Ossetic* XAECYN, daryn (A), (A); *Pashto* LAREL (D), (D); *Wakhi* WUDER (A), (A); *Yaghnobi* dor- (A), (A); *Old Persian* dārayatīy (A), (A); *Khotanese Saka* dar- (A), (A); *Waziri* NIWEL, SOTEL (C), (C);
- horn:** *Sogdian* qrn' (H), (H); *Ormuri* šāx (F), (F); *Larestani* šāx (E), (E); *Avestan* sruīie (du.) (A), (A); *Kurdish* qiloç (B), (B); *Persian* shākh (C), (C); *Kumzari* qarn ( < Ar) (D), (D); *Yaghnobi* šox (G), (G); *Zazaki* qoç (B), (B); *Shughni* xōx (E), (E);
- how:** *Ossetic* kuyd (A), (A); *Pashto* CENGA (A), (A); *Shughni* ar-di, jinow, ca (A), (A); *Balochi* CHACCHO, CHON, CHO (A), (A); *Persian* chetowr (A), (A); *Khotanese Saka* Sū (A), (A); *Avestan* cū (A), (A); *Wakhi* TSERUNG (A), (A); *Yaghnobi* čutte (A), (A); *Zazaki* senīn,čitur (A), (A); *Larestani* cu (A), (A); *Sogdian* c' (A), (A); *Waziri* TSANGRA (A), (A);
- hunt:** *Pashto* SKAR KAVEL (C), (A); *Zazaki* seydnēçir (noun) (B), (B); *Ossetic* cuan kaenyn (D), (C); *Parthian* 'škr (C), (A); *Waziri* SHKORZAN (HUNTER) (C), (A); *Larestani* 'eškāl (noun) (C), (A); *Persian* SHEKAR KARDAN (C), (A); *Shughni* ~γēw (F), (D); *Bandari* eškāl (noun) (C), (A); *Balochi* SHIKAR KHANAGH-A PHA (HUNTING GAME) (A), (A); *Yaghnobi* sayd kun- (E), (E);
- husband:** *Balochi* MARD (C), (C); *Pashto* MERE (C), (C); *Zebaki* māl (C), (C); *Yaghnobi* wir,sarsanka (F), (F); *Waziri* MERE, CHESHTAN (C), (C); *Bandari* šü (D), (D); *Zazaki* mērd(e) (C), (C); *Wakhi* SAUHER (B), (B); *Shughni* čōr (G), (G); *Ossetic* l?g (E), (E); *Avestan* paitiš (A), (A); *Persian* SHOWHAR (D), (D); *Sogdian* wyr (F), (F);
- ice:** *Waziri* KARANG (D), (D); *Pashto* JAX (A), (A); *Balochi* BAWAR (C), (C); *Sogdian* yxn(w) (A), (A); *Yaghnobi* ix (A), (A); *Ossetic* ji? (A), (A); *Ormuri* yax (A), (A); *Wakhi* YUZ (B), (B); *Persian* YAKH (A), (A); *Avestan* aēxəm (A), (A); *Larestani* yax (A), (A); *Parachi* īx (A), (A); *Shughni* yāx (A), (A);
- if:** *Bandari* aga (I), (A); *Zebaki* ke (J), (E); *Sogdian* 'YK (J), (E); *Shughni* aga, agar, ca, tām (N), (A); *Larestani* 'āga (H), (A); *Yaghnobi* agar,magam (M), (A); *Waziri* CHE,

- KE (J), (E); *Old Persian* yadiy (B), (B); *Wakhi* UGER (C), (A); *Ishkashmi* agar, tse (L), (A); *Balochi* AR, AR KI (D), (C); *Persian* AGAR (F), (A); *Khotanese Saka* ka (J), (E); *Ossetic* kuy (J), (E); *Parachi* agar (K), (A); *Zazaki* eger,-se,wexto ki, sew (ki), gāvā (E), (A); *Pashto* KA (J), (E); *Avestan* yat̕ (A), (A); *Kumzari* kā (G), (D);
- in:** *Persian* DAR (C), (C); *Zebaki* ka,pa,pēZ (H), (H); *Ossetic* -MAE, -y, - AEJ (F), (F); *Ishkashmi* dar,pa,po (G), (G); *Balochi* NIANWAN, LAFA, SARA (A), (A); *Pashto* PE...KI (E), (E); *Waziri* KSHE, PA...KSHE (D), (D); *Zazaki* -di (B), (B); *Shughni* pi,tar (I), (I); *Sogdian* pr (I), (I); *Kumzari* indī (B), (B);
- kill:** *Balochi* KHUSHAGH, KHUSHTA (C), (C); *Larestani* košta (C), (C); *Bandari* koš- (C), (C); *Shughni* kȧγ (C), (C); *Kurdish* kuştin (C), (C); *Parthian* 'wjn (A), (A); *Zazaki* kištiš (C), (C); *Zebaki* Zanum (A), (A); *Khwarezmian* wzn (A), (A); *Waziri* WEZHLEL (D), (D); *Avestan* ǰainti (A), (A); *Persian* KOSHTAN (C), (C); *Ossetic* maryn (E), (E); *Yaghnobi* tuxoy-,kuš- (C), (C); *Middle Persian* 'wzn (A), (A); *Wakhi* SAEI- (B), (B); *Pashto* VAZEL (D), (D); *Old Persian* avaǰantiy (A), (A);
- knee:** *Persian* zanu (A), (A); *Kurdish* çok (B), (B); *Yaghnobi* zonk,zonu (A), (A); *Larestani* zānū (A), (A); *Avestan* žnu- (A), (A); *Ishkashmi* zung (A), (A); *Middle Persian* 'šnwg (A), (A); *Khotanese Saka* ysānū (A), (A); *Sogdian* j(n)wwq (A), (A); *Shughni* zun (A), (A); *Zazaki* sāqe,çok (B), (B); *Parachi* zānū (A), (A); *Ormuri* zānū (A), (A);
- know:** *Sogdian* γϱβ (F), (F); *Persian* DANESTAN (A), (A); *Wakhi* DIS- (B), (B); *Old Persian* dānātiy (xšnā-) (A), (A); *Ossetic* faesmaerun, zonym (A), (A); *Yaghnobi* Giriv (F), (F); *Pashto* POHEDEL (E), (E); *Waziri* KHABAR, MOLIM (KNOWN) (D), (D); *Parthian* z'n (A), (A); *Bandari* dānesten (A), (A); *Khotanese Saka* paysān- (A), (A); *Kurdish* zanîn (A), (A); *Avestan* zanāiti, vaēda (A), (A); *Balochi* SAHIH BIAGH (C), (C); *Middle Persian* d'n- (A), (A); *Zazaki* zānāyiš (A), (A); *Shughni* fā m, wizān (A), (A);
- lake:** *Balochi* DHAND (C), (C); *Avestan* vairiš (A), (A); *Ormuri* daṇḍ (H), (H); *Zazaki* dol,gol (D), (D); *Persian* ??, DARYACHE (E), (E); *Ossetic* cad (G), (G); *Middle Persian* dry'b (E), (E); *Shughni* qul (D), (D); *Wakhi* ZOI (B), (B); *Pashto* DAND, KUL, DZIHIL, DARJACA, GADIR (F), (F); *Larestani* daryāce (E), (E); *Parachi* hauz (I), (I); *Khotanese Saka* tcāta- (G), (G); *Sogdian* (swmdr) (J), (J); *Parthian* zryh (E), (E);
- laugh:** *Pashto* XANDEL (A), (A); *Shughni* šānd (A), (A); *Ormuri* xan-ak (A), (A); *Parthian* xnd- (A), (A); *Balochi* KHANDAGH, KHANDITHA (A), (A); *Larestani* xana kerda (A), (A); *Middle Persian* xn- (A), (A); *Khotanese Saka* khan- (A), (A); *Parachi* khan- (A), (A); *Ossetic* xudyn (A), (A); *Yaghnobi* xant- (A), (A); *Zazaki* huw(iy)āyiš (B), (B); *Kumzari* khanīdish (pst) (A), (A); *Sogdian* γnt (A), (A); *Wakhi* KUND (A), (A); *Persian* KHANDIDAN (A), (A); *Waziri* KHANDEL (A), (A);
- leaf:** *Ishkashmi* (barg) (F), (A); *Bandari* tāγ (C), (C); *Pashto* PANA (D), (D); *Parachi* pōn (D), (D); *Khotanese Saka* pirā- (G), (G); *Zazaki* velg (A), (A); *Balochi* THAKH (C),

- (C); *Kurdish* p̄el (B), (B); *Persian* BARG (A), (A); *Ossetic* syf (E), (E); *Larestani* barg (A), (A); *Waziri* PONRYE (D), (D); *Shughni* barg,bark,p̄ark (A), (A); *Wakhi* PULC (B), (B); *Sogdian* wrkr (A), (A); *Avestan* var̄aka- (A), (A);
- left:** *Sogdian* pr'γsy (E), (E); *Yaghnobi* čapa (C), (C); *Shughni* čāp (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* haoiiō (A), (A); *Persian* CHAP (B), (B); *Pashto* KIN (C), (C); *Larestani* cap (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* paitištānəm (A), (A); *Yaghnobi* poda,po,ling (A), (A); *Parachi* leng (B), (B); *Sogdian* p'δ (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* LMOSTEL (E), (E); *Yaghnobi* nepid- (H), (H); *Pashto* PREVATELAJ (3 SG.) (F), (F); *Persian* K HABIDAN (DERAZ KASHIDAN) (D), (D); *Shughni* a~γas (I), (I); *Ossetic* xuyccyn (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* PAEDEL (D), (D); *Parthian* jyw- (A), (A); *Avestan* jūuaiti (A), (A); *Old Persian* jīvatiy (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* igaer (A), (A); *Ishkashmi* gōla (G), (C); *Larestani* jegar (A), (A); *Persian* JEGAR (A), (A); *Waziri* YENNA (A), (A); *Shughni* jīgār,θōd (H), (E); *Avestan* yākarə (A), (A); *Kurdish* ceger (D), (A); *Zazaki* qesibā (E), (B); *Balochi* JAGHAR (C), (A); *Wakhi* JIGAR (B), (A); *Khotanese Saka* gyagarrā (A), (A); *Pashto* INA (A), (A); *Parachi* jigar (F), (A);
- long:** *Old Persian* darga (A), (A); *Khotanese Saka* dāra- (A), (A); *Persian* DERAZ, boland (B), (B); *Pashto* UZD (B), (B); *Avestan* darəγō (A), (A); *Ormuri* drāγ (A), (A); *Ossetic* darg' (A), (A); *Wakhi* DEROZ, VERZ (A), (A); *Waziri* WIZHD (B), (B); *Parthian* drG (A), (A); *Shughni* darōz, vō~γž (B), (B); *Zazaki* derg (A), (A); *Sogdian* βrzyy (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* ešpiž (A), (A); *Shughni* murač, sipa~γ (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* šipúš (A), (A); *Avestan* spiša- (A), (A); *Persian* shepesh (A), (A); *Bandari* šoš/šüş (A), (A); *Sogdian* špšh,ršk' (A), (A); *Waziri* SPAZHA (A), (A); *Ormuri* spōi (A), (A); *Kurdish* sipi (A), (A);

**man:** *Khotanese Saka* daha- (E), (E); *Zebaki nar* (A), (A); *Wakhi* MERDINA, DAEI, XULG (B), (B); *Parthian* myrd,nr (A), (A); *Kurdish* mêr (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); *Zazaki* merdim,mêrdek (B), (B); *Persian* mard, ADAM (B), (B); *Middle Persian* myrd,nr (A), (A); *Avestan* vîrō, nā (A), (A); *Ishkashmi* nark (A), (A); *Balochi* MARD, MAR (B), (B); *Bashkardi* mōš [ < martiya] (B), (B); *Parachi* mē?,nar (A), (A); *Old Persian* \*nar- (A), (A);

**many:** *Shughni* bisyōr? (M), (M); *Ormuri* zut (D), (D); *Parachi* γalaba (I), (I); *Kumzari* khaykeh (E), (E); *Larestani* xáyli, ziyād, basi (F), (F); *Balochi* BA Z (C), (C); *Old Persian* parava (A), (A); *Avestan* parauuō (A), (A); *Yaghnobi* morti (K), (K); *Kurdish* gelek, pîr (A), (A); *Waziri* DER (G), (G); *Ossetic* birae (H), (H); *Wakhi* GHUFC (B), (B); *Sogdian* 'ft'r (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* ggūšta (A), (A); *Zazaki* gošt (A), (A); *Middle Persian* pyt (B), (B); *Yaghnobi* yota,gušt (A), (A); *Ishkashmi* pudf (B), (B); *Parachi* γUš (A), (A); *Parthian* pyt (B), (B); *Wakhi* GOST (A), (A); *Kumzari* gōsht (A), (A); *Waziri* GHOSHA, GHESHA (A), (A); *Shughni* gūxt (A), (A); *Balochi* GOZHD (A), (A); *Persian* gusht (A), (A); *Larestani* güšt (A), (A); *Kurdish* gošt (A), (A); *Sogdian* 'pt (B), (B); *Pashto* GVASA (A), (A); *Ormuri* gâka (C), (C);

**moon:** *Kurdish* meh, heyv (A), (A); *Old Persian* \*māhā- (A), (A); *Parachi* mahök (A), (A); *Ishkashmi* mā (A), (A); *Yazgulyam* māst (A), (A); *Middle Persian* m'h (A), (A); *Zazaki* āšmī (B), (B); *Kumzari* maytāwo (A), (A); *Zebaki* ilmēk (C), (C); *Yaghnobi* mahtop (A), (A); *Parthian* m'h (A), (A); *Ormuri* mātau (A), (A); *Shughni* mēst (A), (A); *Sogdian* m'x (A), (A); *Avestan* mā (A), (A); *Khotanese Saka* māstā (A), (A);

**mother:** *Balochi* MATH (A), (A); *Waziri* MOR, MER (A), (A); *Kumzari* mām (A), (A); *Pashto* MOR (A), (A); *Zazaki* mā(y),dād (A), (A); *Parachi* a'ī,mācī (A), (A); *Khotanese Saka* nāni (B), (B); *Persian* madar (A), (A); *Avestan* māta (A), (A); *Shughni* nān (B), (B); *Ossetic* mad (A), (A); *Larestani* nana (B), (B); *Ishkashmi* nān (B), (B); *Sogdian* m't (A), (A); *Bandari* müm (A), (A); *Wakhi* NUN (B), (B); *Yaghnobi* oča (C), (C); *Old Persian* \*mātā (A), (A);

**mountain:** *Old Persian* kaufa (B), (B); *Wakhi* KU(H) (F), (B); *Avestan* gairiš (A), (A); *Kumzari* kō (B), (B); *Waziri* GHAR (A), (A); *Zazaki* ko (B), (B); *Shughni* kū, puxtā, tāk (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,kō (A), (A); *Sogdian* γr- (A), (A); *Khotanese Saka* ggari- (A), (A); *Larestani* kü (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* āha- (A), (A); *Kumzari* kār (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),pōz (I), (I); *Khwarezmian* k'm (H), (H); *Sogdian* kwc' (K), (H); *Yazgulyam* fōts (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* nēm (A), (A); *Parthian* n'm (A), (A); *Ossetic* nom (A), (A); *Yaghnobi* nom (A), (A); *Persian* ESM (B), (B); *Khotanese Saka* nāma- (A), (A); *Middle Persian* n'm (A), (A); *Avestan* nāma (A), (A); *Ormuri* nām,pazak (A), (A); *Wakhi* NONG (A), (A); *Shughni* num (A), (A); *Kurdish* nav (A), (A); *Waziri* NUM (A), (A); *Zazaki* nām(e) (A), (A); *Sogdian* n'm (A), (A); *Larestani* 'esm (C), (C); *Old Persian* nāmā (A), (A);

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

**near:** *Ossetic* XAESTAEG, cyr (C), (C); *Yaghnobi* qarīb (E), (E); *Ishkashmi* dzā̄ (D), (D); *Waziri* NEZDE, TSARMA (B), (B); *Wakhi* NEZD, QERIB, SIS (B), (B); *Pashto* NEZDE (B), (B); *Kumzari* nayzik (B), (B); *Sogdian* β'w (F), (F); *Avestan* asne (A), (A); *Old Persian* ašnaiy (A), (A); *Balochi* NAZI, NAZIKH (B), (B); *Zebaki* jā (D), (D); *Persian* NAZDIK (B), (B); *Larestani* nazik (B), (B); *Bandari* nazik (B), (B); *Zazaki* nezdī (B), (B); *Shughni* qarīb,nazdik (G), (B);

**neck:** *Wakhi* GERDON, MAEYUK (B), (B); *Kurdish* gerden (B), (B); *Yaghnobi* kama ( ), ( ); *Persian* GARDAN (B), (B); *Zazaki* mil,vil ( ), ( ); *Avestan* manaοθri, grīuua (A), (A); *Balochi* GVAR (B), (B); *Larestani* gardū (B), (B); *Kumzari* gurdin (B), (B); *Shughni* gardān,māk (B), (B); *Sogdian* γrd'k (B), (B); *Waziri* GHWORA, MA KANDAI, MAGHZAI ( ), ( ); *Pashto* GARA (B), (B);

**new:** *Sogdian* nwy (A), (A); *Kurdish* nū (A), (A); *Zazaki* newe (A), (A); *Wakhi* TOZA, SEGHD (B), (B); *Ossetic* naeuag (A), (A); *Persian* TAZE (B), (B); *Parachi* nō (A), (A); *Kumzari* nō (A), (A); *Ishkashmi* nawuk (A), (A); *Waziri* NEWAI (A), (A); *Larestani* now, tāza (A), (A); *Shughni* naw,tirák (A), (A); *Pashto* NEVAJ (A), (A); *Khotanese Saka* nūvara (A), (A); *Balochi* NOKH (A), (A); *Ormuri* nūw (A), (A); *Yaghnobi* nava (A), (A); *Avestan* nauuō (A), (A);

**night:** *Khotanese Saka* SSava (A), (A); *Wakhi* NUGHD (B), (B); *Waziri* SHPA (A), (A); *Middle Persian* šb (A), (A); *Kurdish* şev (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* xāb (A), (A); *Avestan* xšap- (A), (A); *Persian* SHAB (A), (A); *Yaghnobi* xšap,šab (A), (A); *Yazgulyam* Šab (A), (A); *Parthian* šb (A), (A); *Khwarezmian* 'xýb (A), (A); *Kumzari* shō (A), (A); *Ossetic* aexcaev (A), (A); *Bandari* šü (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* nōkharet (A), (A); *Wakhi* MIS (B), (B); *Shughni* nēž (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); *Waziri* PEZA, WARBIZ, WARSAK (H), (H); *Parachi* damâγ,nēšt (I), (I); *Zazaki* pirnike (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* nōit (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* 'wcnny (E), (E); *Khotanese Saka* ysaMgara- (C), (C); *Waziri* ZOR (C), (C); *Avestan* hanō (A), (A); *Shughni* kīnā,pīr (D), (D); *Persian* PIR (B), (B); *Larestani* pir (B), (B); *Wakhi* KONA, XAEIYAR ( ), ( ); *Zazaki* pīl,ixtiyār (B), (B); *Ormuri* zāl (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* aēuuō (A), (A); *Wakhi* YI, I, YIU (A), (A); *Zebaki* wok (A), (A); *Ishkashmi* wak (A), (A); *Old Persian* aiva (A), (A); *Persian* yek, ???? (A), (A); *Waziri* YO (A), (A); *Kurdish* yek, êk (A), (A); *Ossetic* iu (A), (A); *Yaghnobi* yak (A), (A); *Pashto* JAV (A), (A); *Khotanese Saka* śsau- (B), (B); *Yazgulyam* wōG (A), (A); *Shughni* yak,yīw (A), (A); *Bandari* ya (A), (A); *Ormuri* šē (A), (A); *Bactrian* ywg (A), (A);
- other:** *Avestan* ainiiō (A), (A); *Zazaki* bīn (D), (D); *Khwarezmian* 'ny (A), (A); *Yaghnobi* áne,axiš (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* īnsān,merdim (A), (A); *Persian* SHAKHS (C), (C); *Shughni* čōr, nafar (G), (G); *Sogdian* 'γryw(h) (F), (F); *Old Persian* martiya (A), (A); *Larestani* 'ādam (D), (D); *Balochi* KHAS (B), (B); *Pashto* KAS, TAN (E), (E); *Avestan* mašiiākō, mašiiō (A), (A); *Kurdish* meriv, mirov (A), (A);
- play:** *Persian* BAZI KARDAN (C), (C); *Yaghnobi* bozi (noun) (E), (E); *Shughni* bōzi čidōw (G), (G); *Kumzari* baz gu'dish (pst) (C), (C); *Zazaki* kāy (noun) (B), (B); *Balochi* LEV KHANAGH (A), (A); *Middle Persian* w'c (C), (C); *Sogdian* k't'k (F), (F); *Pashto*



- BAZI KAVEL (E), (E); *Waziri* MAZSHILEDEL (D), (D); *Bandari* gāzī (noun) (C), (C); *Ossetic* x'azyn (B), (B);
- pull:** *Zazaki* āntiš (D), (D); *Middle Persian* krš- (B), (B); *Pashto* KSEL, ISTELE (B), (B); *Persian* KASHIDAN (B), (B); *Shughni* tāž (A), (A); *Yaghnobi* čumf-,fuzzon-,sekill- (G), (G); *Waziri* WUKSHEL, KSHEL (B), (B); *Larestani* kašeda (B), (B); *Wakhi* XUS- (B), (B); *Sogdian* 'wst't (H), (H); *Avestan* θanjaiiete (A), (A); *Balochi* CHIKAGH, CHIKITHA (C), (C); *Ossetic* LASYN, ivazyn (F), (F);
- push:** *Ossetic* ycxojyn (F), (F); *Waziri* PORI WAHEL (E), (E); *Persian* HOL DADAN (C), (C); *Balochi* TELAN (SB.) (B), (B); *Larestani* loo dada (D), (D); *Shughni* barγā Dedow (G), (G); *Pashto* PORI WAHEL (E), (E); *Wakhi* SUKE DI-, TECUV- (A), (A);
- rain:** *Sogdian* w'r (A), (A); *Pashto* BARAN UREZI (A), (A); *Ishkashmi* urnaduk (B), (B); *Parthian* w'r (A), (A); *Balochi* GWARAGH, GWARTA (A), (A); *Avestan* vār- (A), (A); *Ossetic* uaryn, K"AEVDA (A), (A); *Zazaki* vārān,yāγer (A), (A); *Yaghnobi* borón (A), (A); *Kumzari* bāram (noun) (A), (A); *Larestani* baru baredan (A), (A); *Shughni* bōrūn (A), (A); *Kurdish* baran (A), (A); *Waziri* WAREDEL (A), (A); *Parachi* auγār,wāš- (A), (A); *Persian* BARIDAN (A), (A); *Middle Persian* w'r (A), (A);
- red:** *Zazaki* sūr (B), (B); *Avestan* raoiditō (A), (A); *Shughni* lōlā,rūšt (H), (H); *Waziri* SIR (B), (B); *Khotanese Saka* ysarūna- (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* krm'yr (F), (F); *Wakhi* SEKR (B), (B); *Parachi* surkhō,surku (D), (B); *Ishkashmi* surx (E), (B); *Persian* SORKH (B), (B);
- right:** *Zazaki* doγrī,rāšt (A), (A); *Old Persian* rāsta (A), (A); *Balochi* RAST (A), (A); *Larestani* rāss (A), (A); *Wakhi* WERTS, DURUST, BUF, BEROBER (B), (B); *Sogdian* dšn (F), (F); *Ormuri* xwarinc (E), (E); *Avestan* ərašuuō (A), (A); *Ossetic* rast (D), (A); *Khotanese Saka* rraSTa- (A), (A); *Persian* DOROST (B), (B); *Waziri* SAHI (C), (C); *Shughni* dirust, māqūl (G), (G); *Pashto* SAM (C), (C);
- rightside:** *Middle Persian* dšn (F), (F); *Shughni* rōst, xēz (H), (H); *Avestan* dašinō (A), (A); *Khotanese Saka* hvarandau (G), (G); *Parthian* dšn (F), (F); *Sogdian* xw'r'nt (E), (E); *Balochi* RAST (C), (C); *Waziri* SHAI (D), (D); *Zazaki* rāšt (C), (C); *Ossetic* RAXIZ, rast (C), (C); *Wakhi* WURZGE (B), (B); *Pashto* SAJ (D), (D); *Persian* RAST (C), (C);
- river:** *Parthian* rwd (A), (A); *Old Persian* rauta (A), (A); *Khotanese Saka* nā,nātā (H), (H); *Zazaki* ro(y),lā,čem (A), (A); *Shughni* daryō (I), (I); *Persian* DARYA (D), (B); *Wakhi* DERIO (B), (B); *Sogdian* rwt (A), (A); *Ormuri* wōkxāna (G), (G); *Waziri* DARYOB, TOI (E), (B); *Avestan* rauua, θraotō (A), (A); *Yaghnobi* rawt,op,dariyo,nahr (A), (A); *Ossetic* DON, caeugaedon (F), (F); *Larestani* rüdxuna, big (a large one) (A), (A); *Pashto* RUD, SIND (A), (A); *Middle Persian* rwd (A), (A); *Bandari* rüxāna (A), (A); *Balochi* DIRA (C), (B);

- road:** *Parthian* r'h (C), (C); *Parachi* panân (A), (A); *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* rāh (C), (C); *Waziri* LYAR (C), (C); *Kurdish* rê (C), (C); *Pashto* LAR (C), (C); *Wakhi* VEDEK (B), (B); *Larestani* jādā, rā (C), (C); *Ormuri* lār, rār (C), (C); *Yaghnobi* rot, rah (C), (C); *Avestan* pantā, aδβa (A), (A); *Old Persian* \*paθi- (A), (A); *Sogdian* r'θ (C), (C); *Ossetic* faendag (A), (A); *Persian* RAH (C), (C); *Balochi* DAG (B), (B);
- root:** *Wakhi* WIUX (D), (D); *Ossetic* BYN, uidag ( ), ( ); *Shughni* bex, wex, wiyeš (D), (D); *Larestani* riša (A), (A); *Sogdian* wyx (D), (D); *Balochi* PAR (B), (B); *Kumzari* irq (C), (C); *Ormuri* bēx (D), (D); *Pashto* BEX, RISA (D), (D); *Zazaki* riče (A), (A); *Persian* rishe (A), (A); *Avestan* varəšaǰiš (A), (A); *Parachi* bīx/γīx, kōrdi (D), (D); *Kurdish* reh (A), (A); *Yaghnobi* riša (E), (E); *Waziri* BEKH, WEKH (D), (D);
- rope:** *Ishkashmi* vuš (C), (C); *Shughni* bānd, sarbānd (E), (E); *Yaghnobi* wíta (D), (D); *Balochi* REZ (B), (B); *Zazaki* resen (B), (B); *Waziri* PERAI ( ), ( ); *Pashto* PERAJ ( ), ( ); *Larestani* band, resmu (C), (C); *Wakhi* TUNOV, DEROWI, NUS, SIVEN (A), (A); *Khwarezmian* rsŷn (B), (B); *Zebaki* wāš (C), (C); *Kumzari* bayn (C), (C); *Parachi* bāš (C), (C); *Persian* TANAB (A), (A); *Bandari* risamün (B), (B); *Ossetic* baendaen (C), (C);
- rotten:** *Sogdian* pwtqy (A), (A); *Ormuri* šri-būk (D), (D); *Waziri* WROST (D), (D); *Persian* PUSIDE (C), (C); *Khotanese Saka* haMbūta- (A), (A); *Ossetic* aembyd (A), (A); *Yaghnobi* pūta (A), (A); *Avestan* pūtō (A), (A); *Pashto* VROST (D), (D); *Balochi* GALAGH (B), (B); *Wakhi* PITK (A), (A);
- round:** *Larestani* gerd (B), (B); *Kurdish* girover (B), (B); *Zazaki* tekele (C), (C); *Yaghnobi* lunda, kursak (D), (D); *Sogdian* 'skwrnk (E), (E); *Avestan* skarənō (A), (A); *Shughni* žarn, žurn (F), (F);
- rub:** *Wakhi* MAND, SUX (A), (A); *Zazaki* sāvitiš (B), (B); *Larestani* māleda mošta (A), (A); *Waziri* MASHEL (A), (A); *Parachi* astar (B), (B); *Sogdian* 'nsy- (B), (B); *Ormuri* say- (B), (B); *Pashto* MUSEL (A), (A); *Ossetic* xafŷn (C), (C); *Balochi* MALAGH, MALITHA, MALTHA (A), (A); *Khotanese Saka* dar- (D), (D); *Persian* MALIDAN ( ), ( );
- salt:** *Ishkashmi* nāmūlGak (A), (A); *Persian* NAMAK (A), (A); *Khotanese Saka* namva- (A), (A); *Khwarezmian* šwr (F), (F); *Ormuri* nimēk (H), (A); *Balochi* WHADH, WAHADH (B), (B); *Pashto* MALGA (A), (A); *Shughni* namák (G), (A); *Parachi* namâ (A), (A); *Kumzari* khūwah (D), (D); *Yaghnobi* namak (A), (A); *Sogdian* nm'δk (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, māsa, jong (finer one), lamr (finest one) (D), (D); *Bandari* γiyām (F), (F); *Shughni* šōš, rēg (G), (G); *Khotanese Saka* syatā (G), (G); *Persian* SHEN (D),

- (D); *Waziri* SHEGGA (G), (G); *Khwarezmian* cγ (G), (G); *Balochi* REKH (B), (B); *Kumzari* dīrī (E), (E); *Ormuri* (rēg) (I), (B); *Wakhi* LUWORC (A), (A); *Kurdish* rik, qum (B), (B); *Ishkashmi* rēg (J), (B); *Pashto* SEGA, RIG (G), (G); *Yaghnobi* reg (K), (B); *Parachi* seγa (G), (G); *Ossetic* yzmic (H), (H); *Zazaki* qūm (C), (C);
- say:** *Zazaki* vātīš (E), (E); *Zebaki* GeZ- (F), (F); *Balochi* GUSHAGH, GUSTA, GWASHAGH, GWASHTA (D), (D); *Avestan* aoخته, mraoitī (A), (A); *Ishkashmi* GeZ- (F), (F); *Waziri* WEYEL (E), (E); *Ossetic* dzuryn ( ), ( ); *Kurdish* gotin (D), (D); *Bashkardi* go (D), (D); *Persian* GOFTAN (D), (D); *Bandari* goften (D), (D); *Pashto* VAJEL (E), (E); *Old Persian* θātiy (θah-) (B), (B); *Wakhi* XAN (C), (C); *Ormuri* γuš (D), (D); *Shughni* gāp dedōw (D), (D); *Yaghnobi* wov- (G), (G); *Larestani* gota (D), (D);
- scratch:** *Wakhi* DRUP-, CUNGOL DI- (A), (A); *Zazaki* wirīnāyiš (C), (C); *Persian* KHARIDAN (B), (B); *Yaghnobi* ručon- (E), (E); *Balochi* KHARAGH, KHARITH (B), (B); *Shughni* čangōl dedow (F), (F); *Pashto* GEREDDEL (D), (D); *Waziri* GARAWEL (D), (D); *Larestani* xaraneda (B), (B);
- sea:** *Bandari* deryā (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* dendzyz (F), (F); *Kumzari* derīyō (A), (A); *Larestani* daryā (A), (A); *Avestan* zraiiō (A), (A); *Sogdian* (swmdr-) (G), (G); *Middle Persian* dry’b (A), (A); *Pashto* BAHAR (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* dītiš (B), (B); *Waziri* KATEL, LIDEL (C), (C); *Yaghnobi* wen- (A), (A); *Avestan* vaēnaiti (A), (A); *Kurdish* dītin (B), (B); *Khotanese Saka* tcäs (D), (D); *Ossetic* uynyn (A), (A);
- seed:** *Avestan* taoxma, čiθrəm (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* teGm (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, tuxm (A), (A); *Balochi* bidz (B), (B); *Kurdish* bizir, tov (B), (B); *Khotanese Saka* ttīma- (A), (A);
- sew:** *Zazaki* derzāyiš (C), (C); *Shughni* pāmi (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); *Wakhi* DREV- (A), (A);
- sharp:** *Waziri* TERA (A), (A); *Persian* TIZ (A), (A); *Wakhi* TEGHD, TIZ (A), (A); *Avestan* tiγrō (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* tigrā (A),

- (A); *Ishkashmi* tēz (C), (A); *Yaghnobi* tez (D), (A); *Sogdian* sqrb, trγ (A), (A); *Parachi* tēz (B), (A); *Pashto* TERE (A), (A); *Middle Persian* tyj (A), (A);
- short:** *Waziri* LAND (E), (E); *Zazaki* kilm (D), (D); *Pashto* MUXTASAR, LAND (E), (E); *Persian* kutah (B), (B); *Sogdian* mwrzk (A), (A); *Balochi* GWAND (C), (C); *Avestan* mərəzu- (A), (A); *Parachi* luṇḍu (H), (H); *Khotanese Saka* mulysga- (A), (A); *Ormuri* laṇḍ (G), (G); *Parthian* qmbyg (J), (J); *Shughni* kat, kut (I), (I); *Yaghnobi* kaltá (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* šarvand zaden (D), (D); *Ossetic* zaryn (G), (G); *Yaghnobi* žoy, xofizi kun- (I), (I); *Persian* AVAZ KHANDAN (C), (C); *Wakhi* BAEIT XAN- (A), (A); *Balochi* GUSHAGH, GUSHTA, GWASHTA (B), (B); *Zebaki* GēZāk (H), (H); *Shughni* luv (K), (K);
- sit:** *Larestani* šessa (C), (C); *Persian* NESHASTAN (C), (C); *Shughni* niθ (B), (B); *Yaghnobi* nid- (B), (B); *Kurdish* danîşîn, rûniştin (C), (C); *Bandari* nešten (C), (C); *Ishkashmi* nēd- (B), (B); *Parthian* nšyy- (B), (B); *Ossetic* badyn (E), (E); *Middle Persian* nšyy- (B), (B); *Zebaki* nīd- (B), (B); *Bashkardi* nen (C), (C); *Pashto* KSENASTELE (D), (D); *Avestan* āste (A), (A); *Sogdian* nst- (B), (B); *Wakhi* NEZD (B), (B); *Waziri* KSHENAWEL, NOSTAI, PAND, (SITTING) (D), (D); *Balochi* NINDAGH, NISHTA (B), (B); *Khotanese Saka* Nāšta (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* pūst (H), (B); *Shughni* pust (J), (B); *Ossetic* carm (C), (C); *Persian* pust (E), (B); *Khwarezmian* crm (C), (C); *Kumzari* pōst (F), (B); *Khotanese Saka* kangā-, chala, ttanā (D), (D); *Middle Persian* pwst, crm (C), (B); *Balochi* PHOST (C), (B); *Yaghnobi* pust (I), (B); *Avestan* suri (A), (A);
- sky:** *Wakhi* OSMON (A), (A); *Yaghnobi* osmon (A), (A); *Pashto* ASMAN (A), (A); *Avestan* asma (diiau- Daevic) (A), (A); *Bashkardi* yāhmōn (A), (A); *Yazgulyam* asmīn (A), (A); *Shughni* ōsmun (A), (A); *Waziri* ASMON (A), (A); *Zazaki* āzmīn (A), (A); *Persian* ASEMAN (A), (A); *Parachi* āγeš < Indic? (C), (C); *Ishkashmi* āsmān (A), (A); *Middle Persian* (')sm'n (A), (A); *Old Persian* asmā (A), (A); *Balochi* ARSH, AZMAN (A), (A); *Kumzari* asmay'nō (A), (A); *Ormuri* āsmān, falak (A), (A); *Parthian* (')sm'n (A), (A); *Sogdian* ('k'c) (D), (D); *Ossetic* arv (B), (B);
- sleep:** *Ormuri* xau (A), (A); *Zazaki* rā-kewtiš, witiš (A), (A); *Shughni* xōfc (A), (A); *Yaghnobi* ufš, nepid (A), (A); *Khotanese Saka* ūm (G), (G); *Sogdian* γwβn- (A), (A); *Avestan* xvafsaite (A), (A); *Khwarezmian* 'zmxs (F), (F); *Waziri* KHEB (A), (A); *Persian* KHABIDAN (A), (A); *Middle Persian* hwps- (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

- (C), (C); *Bandari* xāften (A), (A); *Parthian* xwsp- (A), (A); *Kumzari* khuwōw (imp sg) (A), (A);
- small:** *Ishkashmi* cutōkok (I), (I); *Zazaki* werdī,qiž,šenik (E), (E); *Kumzari* chik (D), (D); *Khotanese Saka* jseiNa- (K), (K); *Larestani* kaydü (E), (E); *Waziri* KAM, KAMKAI, WRIKAI WOR (D), (D); *Shughni* -buc,žal ( ), ( ); *Ormuri* zārī (G), (G); *Parachi* čimō (H), (H); *Wakhi* ZUQ, ZUQIQ, ZUQULAEI (B), (B); *Kurdish* biçûk (D), (D); *Sogdian* kβn'kk (D), (D); *Balochi* KIK (C), (C); *Avestan* kasuš (A), (A); *Zebaki* cuT (I), (I); *Persian* kuchek (D), (D); *Middle Persian* qwdk (E), (E); *Pashto* KUCNAJ, VOR (D), (D); *Ossetic* cysyl, gycyl (F), (F); *Yaghnobi* pulla,maydaak (J), (J);
- smell:** *Balochi* BO GIRAGH (A), (A); *Shughni* buy čidōw (A), (A); *Larestani* bü kerda (A), (A); *Persian* BU KARDAN (BU SHENIDAN) (A), (A); *Zazaki* boy k [sic?] (A), (A); *Pashto* HYS KAVEL (B), (B);
- smoke:** *Pashto* DUD (A), (A); *Wakhi* DIT (A), (A); *Yaghnobi* payst,dud (B), (B); *Ishkashmi* dit (A), (A); *Persian* DUD (A), (A); *Larestani* düd (A), (A); *Shughni* δud (A), (A); *Ormuri* dūd (A), (A); *Sogdian* pzt (B), (B); *Khotanese Saka* dumä (A), (A); *Yazgulyam* ?ād (A), (A); *Parachi* dhī (A), (A); *Kurdish* dûkel, dûxan (A), (A); *Zazaki* dūy (A), (A); *Middle Persian* dwd (A), (A); *Ossetic* faezdaeg, x'uaecae (B), (B); *Balochi* DUHON (A), (A); *Parthian* dwd (A), (A); *Kumzari* dūr (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); *Persian* SAF (C), (C); *Yaghnobi* lixna,hamwor (A), (A); *Wakhi* LUS, HUNWOR (A), (A); *Pashto* SAF, MUSTAVI, HAVAR (E), (E); *Shughni* amwor, daxt (A), (A);
- snake:** *Sogdian* kyrm (E), (E); *Khotanese Saka* śsaysda- (G), (G); *Waziri* MANGER (C), (C); *Wakhi* FUKS (B), (B); *Ishkashmi* voks (F), (F); *Persian* mar (C), (C); *Zazaki* mār (C), (C); *Larestani* mār (C), (C); *Balochi* MAR (C), (C); *Ormuri* mār (C), (C); *Avestan* ažiš (A), (A); *Pashto* MAR (C), (C); *Yaghnobi* mor (C), (C); *Parachi* haždâr,kirm (E), (E); *Shughni* divūs, mōr (H), (H); *Ossetic* kalm (D), (D); *Kumzari* mār (C), (C);
- snow:** *Larestani* vafr, barf [latter < Pers?] (A), (A); *Shughni* dinyō,žini (D), (D); *Ormuri* γōš (A), (A); *Persian* BARF (A), (A); *Parthian* wfr,wpr (A), (A); *Balochi* BAWAR (A), (A); *Avestan* vafra- (A), (A); *Ishkashmi* varf (A), (A); *Parachi* γarp (A), (A); *Ossetic* mit (C), (C); *Waziri* WOVRA (A), (A); *Yaghnobi* wafr (A), (A); *Middle Persian* wfr,wpr (A), (A); *Sogdian* wfry' (A), (A); *Wakhi* ZEM (B), (B); *Zazaki* vewr (A), (A); *Pashto* VAVRA (A), (A);
- some:** *Pashto* CO, CE KADR (D), (D); *Yaghnobi* čof,čandin (E), (E); *Persian* BA'ZI (E), (E); *Shughni* ač, bāzi, čand (H), (H); *Balochi* KHARD-E (C), (C); *Old Persian* \*kaičiy (A), (A); *Ossetic* CYSYL, caldaer (G), (G); *Wakhi* TSUM, TSUMER, KUMD, CIZ (B), (B); *Zazaki* čend,tāy(n) (D), (D); *Waziri* DZENE, TSE (F), (F);

**spit:** *Wakhi* SEX (B), (B); *Shughni* tuf čidōw (G), (G); *Ossetic* tu kaenyn (C), (C); *Sogdian* γw”β (E), (E); *Avestan* spāma- (nm.) (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* darədairaiāt (opt.) (A), (A); *Sogdian* k’β (H), (H); *Persian* NESF KARDAN (D), (D); *Waziri* CHAWEL (E), (E);

**squeeze:** *Pashto* KSEKSEL (D), (D); *Balochi* DABAGH, DABITHA (B), (B); *Persian* FESHAR DADAN (C), (C); *Wakhi* TRUNJ-, FERIL, WEZEM- (A), (A); *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* laeuuyn (E), (E); *Yaghnobi* ušt- (A), (A); *Sogdian* ’wst- (A), (A); *Shughni* wirāfč (F), (F); *Kurdish* rabûn, westîn (B), (B); *Bandari* vustāden (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* xitērž (A), (A); *Kurdish* stêr (A), (A); *Balochi* ISTAR (A), (A); *Zebaki* sitāra (A), (A); *Ishkashmi* struk (A), (A); *Yaghnobi* bildinga, sitora (C), (C); *Sogdian* ’st’r’kt (A), (A); *Middle Persian* ’st’rg (A), (A); *Yazgulyam* Štarāk (A), (A); *Parthian* st’rg (A), (A); *Larestani* setāra [Skjærvø mentions that L has an r > l change] (B), (B); *Khotanese Saka* stāraa- (A), (A); *Kumzari* starg (A), (A); *Avestan* stārō (pl.) (A), (A); *Wakhi* STOR (A), (A); *Ossetic* yst’aly (A), (A); *Ormuri* stirrak (A), (A); *Parachi* estēč (A), (A); *Persian* SETARE (A), (A); *Pashto* STORAJ (A), (A); *Zazaki* estāre (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’β’k (J), (J); *Waziri* LARGAI (F), (F); *Kumzari* bākūr (E), (E); *Wakhi* GHWUZ, SUNG (A), (A); *Khotanese Saka* daula (K), (K); *Zazaki* čiwe, ušir (C), (C); *Ishkashmi* Gufca (H), (H); *Persian* CHUB (D), (D); *Shughni* muqōbā (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* dadāye (M), (M); *Sogdian* sng (A), (A); *Shughni* sang, žir (A), (A); *Bashkardi* sax (G), (G); *Zazaki* šī (E), (E); *Wakhi* GHAR (B), (B); *Ossetic* dur (J), (J); *Parachi* gir (B), (B); *Yaghnobi* sañk (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, aθanga- (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); *Wakhi* ROST (A), (A); *Balochi* SIDHA (B), (B); *Ormuri* rāst (A), (A); *Parthian* r’št (A), (A); *Avestan* ərəzuš (A), (A); *Shughni* čuk, caq (E), (E); *Kumzari* aydil (C), (C); *Middle Persian* r’st (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* RUDEL (C), (C); *Waziri* RAVDEL (C), (C);
- sun:** *Middle Persian* xwr (A), (A); *Balochi* RO, ROSH (B), (B); *Kurdish* roj, xor (B), (B); *Wakhi* OFTOB, YIR (A), (A); *Shughni* xir, ōftōb (A), (A); *Waziri* LMER, MYER (D), (D); *Zebaki* ōrmōzd (G), (G); *Khwarezmian* ’xŷr (A), (A); *Ishkashmi* rēmuz (F), (F); *Pashto* LMAR (D), (D); *Larestani* ‘aftow, xoršid (C), (C); *Yazgulyam* xvōr (A), (A); *Khotanese Saka* urmaysde- (G), (G); *Ormuri* meš (E), (E); *Persian* AFTAB, KHORSHID (C), (C); *Kumzari* intāfō’ (C), (C); *Sogdian* γwr (A), (A); *Zazaki* rož (B), (B); *Yaghnobi* xur,aftob (A), (A); *Avestan* huuarə (A), (A); *Ossetic* xur (A), (A); *Parthian* myhr (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* δm’s (B), (B); *Wakhi* PEDEMES- (B), (B); *Larestani* bād kerda, ‘a barā ‘onda (D), (D); *Ossetic* RAESIJYN, dymyn (B), (B); *Khotanese Saka* nar- (F), (F); *Shughni* waram čidow (G), (G);
- swim:** *Persian* SHENA KARDAN (A), (A); *Shughni* wāz,xinow (A), (A); *Pashto* LAMBO VAHEL (C), (C); *Balochi* TARAGH, TARATHA, THAHARAGH, THAHARTHA (B), (B); *Kumzari* shinaw gi’dish (pst) (A), (A); *Larestani* šenaw kerda (A), (A); *Kurdish* ajnaberī listin, mele kirin (A), (A); *Wakhi* USINAWERI TSER, QELOC XUS-, WEZAN DĪ- (A), (A); *Parachi* aubâzī- (A), (A); *Zazaki* āzne (A), (A); *Sogdian* fsn’y- (A), (A); *Ossetic* lenk kaenyn, nakae kaenyn (D), (D); *Waziri* LAMBEYA (SWIMMING) (C), (C);
- tail:** *Pashto* LAKEJ (D), (D); *Parachi* dōγund (A), (A); *Khotanese Saka* dumaa- (A), (A); *Wakhi* KICIKAM (B), (B); *Balochi* DUMB (A), (A); *Khwarezmian* δwm (A), (A); *Yaghnobi* duym,dumba (A), (A); *Avestan* dumaa- (A), (A); *Kurdish* boç, terī (C), (C); *Persian* dom (A), (A); *Ishkashmi* dumb (A), (A); *Zazaki* poč (C), (C); *Sogdian* δwm (A), (A); *Larestani* dom (A), (A); *Ossetic* dymaeg (A), (A); *Waziri* LAKAI (D), (D); *Bandari* dom (A), (A); *Shughni* dum (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* ucy (B), (B);

- Balochi* AN (B), (B); *Old Persian* ava (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(či) (F), (F); *Old Persian* avadā (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* estörö (G), (G); *Balochi* THULAR (B), (B); *Khotanese Saka* dara- (I), (I); *Pashto* PEND, ZAXIM (E), (E); *Yaghnobi* Gafs,lula,farbi (H), (H); *Ossetic* ystavd (F), (F); *Wakhi* BAJ (A), (A); *Larestani* koloft, got (C), (C); *Shughni* farbiyaki (H), (H);
- thin:** *Larestani* nāzok, tanok (D), (D); *Khotanese Saka* ttaMga- (F), (F); *Balochi* LAGHAR (B), (B); *Shughni* birik, bōrik, nōzuk, tānuk (F), (F); *Ishkashmi* tanuk (F), (F); *Zazaki* bāri?ek (C), (C); *Yaghnobi* tunuk (F), (F); *Persian* BARIK, NAZOK (D), (D); *Sogdian* γymr (G), (G); *Ossetic* taenaeg (F), (F); *Wakhi* SENOR, SENUF, TENUK, BIRIK, XEROB (A), (A); *Pashto* NARAJ (E), (E); *Waziri* NARAI (E), (E);
- think:** *Parthian* nywr- (H), (H); *Old Persian* maniyataiy (A), (A); *Balochi* ZANAGH, ZANTHA (B), (B); *Shughni* fikri čidow, xayol čidow (I), (I); *Bandari* fekr akerden (D), (D); *Ossetic* x'uydy kaenyn (F), (F); *Sogdian* šm'r (G), (G); *Pashto* FIKR KAVEL, GUMAN KAVEL (E), (E); *Avestan* mainiiete (A), (A); *Persian* FEKR KARDAN (C), (C);
- this:** *Persian* IN (A), (A); *Bactrian* ei(i)o (A), (A); *Avestan* aēm (A), (A); *Balochi* HAM-ESH (A), (A); *Zebaki* am (A), (A); *Waziri* DAI, DA, DAGHA (B), (B); *Kurdish* ev (A), (A); *Kumzari* iyah (A), (A); *Sogdian* 'yδ (A), (A); *Pashto* DA, DAGA (B), (B); *Khotanese Saka* tta- (E), (E); *Larestani* 'éde, 'e (B), (B); *Ishkashmi* nakavī (C), (C); *Wakhi* YEM, HAEIEM (A), (A); *Yaghnobi* iš (D), (D); *Parachi* ē (A), (A); *Old Persian* hauv, iyam (A), (A);
- thou:** *Wakhi* TU (A), (A); *Zebaki* tō (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* tō ( ), ( ); *Zazaki* ti (A), (A); *Khotanese Saka* tha (A), (A); *Avestan* tūm, θβam (A), (A); *Pashto* TE (A), (A); *Khwarezmian* 'wtk (A), (A); *Old Persian* tuvam, θuvām (A), (A); *Waziri* TE (A), (A); *Shughni* tu, tōθ (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* šō (A), (A); *Pashto* DRE (A), (A); *Ossetic* aertae (A), (A); *Parachi* ši,šu (A), (A); *Persian* SE (A), (A); *Ishkashmi* rūi (A), (A); *Zazaki* hīrē (A), (A); *Yazgulyam* tsoi (A), (A); *Yaghnobi* tiray,se (A), (A); *Avestan* θrāiiō (A), (A); *Sogdian* šy (A), (A); *Parthian* hry (A), (A); *Khotanese Saka* drai (A), (A); *Shughni* ara,aray,sē (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* ahīyatiy (A), (A); *Zazaki* eštiš,vistiš (D), (D); *Avestan* a?hiieiti, spaieiti (A), (A); *Pashto* ACAVEL, GURZAVEL (H), (H); *Khotanese Saka* ah (K), (A); *Yaghnobi* wid-, pārtow- (J), (F); *Balochi* PHIRENAGH, PHIRENTHA (C), (C); *Shughni* sikawak dedow (L), (J); *Ossetic* aepparyn (I), (I); *Larestani* besa (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, hišāiia (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* vīnd (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* hizua (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); *Shughni* lavz,ziv (A), (A); *Sogdian* zb'q (A), (A); *Persian* zaban (A), (A); *Wakhi* ZIK (A), (A); *Parthian* 'zb'n (A), (A);

**tooth:** *Khotanese Saka* dandaa- (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* GHWOSH, GHOSH (B), (B); *Ishkashmi* dānd (A), (A); *Kurdish* didan, diran (A), (A); *Persian* dandan (A), (A); *Avestan* dantānō (pl.), dātā- ( ), ( ); *Ossetic* daendag (A), (A); *Pashto* GAS (B), (B); *Parachi* danān (A), (A); *Ormuri* gišī (B), (B); *Shughni* dindūn (A), (A);

**tree:** *Balochi* DRASHK (B), (B); *Zazaki* dār (B), (B); *Zebaki* daraxt (F), (F); *Khotanese Saka* kišaukā,ttīra- (G), (G); *Ormuri* d(a)raxt (E), (E); *Ossetic* baelac (D), (D); *Pashto* DIRAXT, VENA (A), (A); *Kumzari* shidreh (C), (C); *Shughni* d(i)raxt, dōrg (B), (B); *Persian* derakht (B), (B); *Yaghnobi* daráxt,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* vana (A), (A); *Middle Persian* \*d'rw (B), (B); *Khwarezmian* wn'yk (A), (A); *Sogdian* wn' (A), (A);

**turn:** *Ossetic* yzdaxyn (E), (E); *Old Persian* vartataiy (A), (A); *Sogdian* prn'z (G), (G); *Avestan* uruuisiieiti (A), (A); *Shughni* garDēn (B), (B); *Zazaki* čerx (C), (C); *Yaghnobi* laksón-,tob- (F), (F); *Wakhi* γir(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); *Sogdian* 'δw' (A), (A); *Balochi* DO (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* dōv (A), (A); *Kurdish* du (A), (A); *Parachi* du (A), (A);

**vomit:** *Waziri* KAI (F), (F); *Balochi* UCHALNA (C), (C); *Wakhi* WOQ- (B), (B); *Pashto* KANGI KAVEL (G), (G); *Shughni* cuγ dedōw, pirīθ (I), (I); *Yaghnobi* diriš-,puxoy- (H), (H); *Larestani* qesun kerda (E), (E); *Khotanese Saka* bāma (A), (A); *Persian* ESTEFRAG KARDAN (OQ ZADAN) (D), (D); *Avestan* vam- (A), (A); *Ossetic* wæmyn, TONYN, omyn (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* rā ceda, gešta (F), (F); *Kurdish* çûn (B), (B); *Ossetic* caeuy (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* garēmō, taptō (A), (A); *Pashto* TOD (B), (B); *Sogdian* šw- (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* γrmnd (A), (A); *Larestani* garm (A), (A); *Shughni* gārm (A), (A);

**wash:** *Wakhi* WUZDI- (A), (A); *Shughni* zini (A), (A); *Avestan* naēnižaiti, snaiieiti (A), (A); *Zazaki* šitiš (B), (B); *Sogdian* sn'y (A), (A); *Larestani* šossa (B), (B); *Yaghnobi* sinoy (A), (A); *Balochi* SHODHAGH, SHUSTA (B), (B); *Ossetic* aexsyn (A), (A); *Kumzari* chōr (sg) (C), (C); *Middle Persian* sn'y (A), (A); *Pashto* MINDZEL (D), (D); *Bandari* šüšten (B), (B); *Waziri* WINZEL (D), (D); *Persian* SHOSTAN (B), (B);

**water:** *Bashkardi* SBš. yāp, NBš. yā/āu (A), (A); *Khotanese Saka* ūtcā (B), (B); *Wakhi* YUPK (A), (A); *Balochi* AF (A), (A); *Parachi* âwə (A), (A); *Yaghnobi* op (A), (A); *Avestan* āfš (A), (A); *Larestani* 'aw (A), (A); *Pashto* OBE (A), (A); *Persian* AB (A), (A); *Ossetic* don (B), (B); *Zazaki* āw (A), (A); *Bandari* hü (A), (A); *Sogdian* 'p'p (A), (A); *Bactrian* y'β (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* mâ (A), (A); *Bashkardi* yamah (A), (A); *Zazaki* mā (A), (A); *Persian* ma (A), (A); *Wakhi* SUK (B), (B); *Kurdish* em, me (A), (A); *Balochi* MA (A), (A); *Avestan* ahma, vaēm (A), (A); *Bandari* mā (A), (A); *Old Persian* vayam (A), (A); *Pashto* MUZ (A), (A); *Shughni* māš (A), (A); *Sogdian* m’x (A), (A); *Waziri* MIZH (A), (A); *Ossetic* max, maxtae (A), (A); *Larestani* ‘amā [not cognate to Pers MA, MA];

**wet:** *Shughni* čālīn, nāmbdōr, xist (A), (A); *Bandari* tar (C), (C); *Ormuri* šūr (G), (G); *Larestani* tal (C), (C); *Persian* TAR (C), (C); *Avestan* naptō (A), (A); *Balochi* THAR (C), (C); *Parachi* phyö (H), (H); *Yaghnobi* tann,obnoka,tar (C), (C); *Ishkashmi* šuhluk (G), (G); *Zazaki* hī (D), (D); *Ossetic* xuylydz (F), (F); *Pashto* XIST, LUND (E), (E); *Wakhi* XUSC (B), (B); *Waziri* LIMD, TOND (E), (E); *Sogdian* nβtc’ (A), (A);

**what:** *Avestan* kaṭ (A), (A); *Sogdian* cw (A), (A); *Waziri* KIM, TSE (A), (A); *Balochi* KITHAN (A), (A); *Parthian* cy (A), (A); *Ishkashmi* cīz (A), (A); *Khotanese Saka* aštū,ca (A), (A); *Yaghnobi* čo (A), (A); *Parachi* če (A), (A); *Larestani* ce, ci (A), (A); *Bandari* či (A), (A); *Middle Persian* cy (A), (A); *Shughni* ca (A), (A); *Kurdish* çı (A), (A); *Pashto* CE (A), (A); *Zebaki* tsīz (A), (A); *Zazaki* čičī (A), (A); *Khwarezmian* c (A), (A); *Persian* che (A), (A); *Wakhi* CIZ, CIZ (A), (A);

**when:** *Wakhi* TSOGHD, TSOGHDER, TSEWUXT (A), (A); *Persian* key (A), (A); *Kumzari* kayī (A), (A); *Pashto* KELA (A), (A); *Balochi* KHADHE (A), (A); *Sogdian* krwδ (A), (A); *Larestani* key (A), (A); *Shughni* ca, ca waxt (B), (B); *Ormuri* ka (A), (A); *Ossetic* kaed, kuy (A), (A); *Yaghnobi* kad (A), (A); *Waziri* KALLA (A), (A); *Khotanese Saka* ku (A), (A); *Avestan* kaḏa (A), (A);

**where:** *Sogdian* ‘kwrδ (A), (A); *Pashto* CIRI, CIRTA (A), (A); *Khotanese Saka* ku (A), (A); *Kumzari* gāyā (A), (A); *Shughni* ar ka, kačad (A), (A); *Ossetic* kaem, kaemyty (A), (A); *Yaghnobi* ku (A), (A); *Wakhi* KUMER, KUMJAEI (A), (A); *Balochi* THANGO (B), (B); *Larestani* koya, ko (A), (A); *Waziri* CHERE (A), (A); *Zazaki* kotī ,kurā,ku,čā (A), (A); *Persian* koja (A), (A); *Avestan* kuθra, kū (A), (A); *Ormuri* gudā (A), (A); *Bandari* kojā (A), (A);

**white:** *Ormuri* spēw (A), (A); *Zazaki* sipī (A), (A); *Waziri* SPIN (A), (A); *Sogdian* ‘sp’yt (A), (A); *Bashkardi* SBš. espīt,NBš. espīr (A), (A); *Ossetic* urs (C), (C); *Pashto* SPIN (A), (A); *Avestan* spaēta-, auriušō (A), (A); *Persian* SEFID (A), (A); *Yaghnobi* sipēta, safēd (A), (A); *Balochi* SWETH (A), (A); *Zebaki* surxūn (E), (E); *Parthian* ‘spyd (A), (A); *Wakhi* ROXUN (B), (B); *Larestani* safid (B), (B); *Kumzari* spīr (A), (A); *Parachi* čaṭō (D), (D); *Khotanese Saka* śśīta- (A), (A); *Middle Persian* ‘spyd (A), (A); *Kurdish* spī (A), (A); *Shughni* safēd (D), (D); *Ishkashmi* safēd (C), (C);

**who:** *Zazaki* kām (A), (A); *Shughni* ar tām (B), (B); *Parachi* kī (A), (A); *Bactrian* kyd (A), (A); *Persian* ki (A), (A); *Bandari* ki (A), (A); *Old Persian* ka (A), (A); *Zebaki* kāi (A), (A); *Yaghnobi* ka,kax (A), (A); *Parthian* ky (A), (A); *Kurdish* kī (A), (A); *Waziri* TSOK (A), (A); *Avestan* kō (A), (A); *Pashto* COK (A), (A); *Sogdian* qy (A), (A); *Balochi*

- KITHAN (A), (A); *Wakhi* KUI (A), (A); *Larestani* ke,ki,ka (A), (A); *Ossetic* ci, citae (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* pərəθuš, paθanō (A), (A); *Yaghnobi* yaxt,pahm (A), (A); *Pashto* PRAX (A), (A); *Waziri* PLAN (A), (A); *Wakhi* FERUX, KESOD (A), (A); *Shughni* daḡt, firōx (B), (B); *Sogdian* pδn’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, nāirīka (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* ĵinī (C), (C); *Ishkashmi* Zānj (C), (C); *Pashto* SEDZA (E), (E); *Shughni* ~γin (C), (C);
- wind:** *Waziri* BOD (A), (A); *Persian* BAD (A), (A); *Larestani* bād (A), (A); *Pashto* BAD (A), (A); *Zazaki* vā (A), (A); *Ossetic* DYMGAE, uad (A), (A); *Avestan* vātō (A), (A); *Yaghnobi* wot,bóda,samól (A), (A); *Sogdian* w’δ (A), (A); *Parachi* dhamān (D), (D); *Wakhi* SEMOL, DUMA (B), (B); *Shughni* šamōl, žuš (E), (E); *Balochi* GO (C), (C);
- wing:** *Parthian* b’zwr (B), (B); *Sogdian* w’z (B), (B); *Avestan* parənəm (A), (A); *Balochi* PHAR (A), (A); *Shughni* pār (A), (A); *Zazaki* peṛ (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* saerfyn (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); *Wakhi* DU (B), (B); *Persian* BA (D), (D); *Avestan* haḡa, maḡ (A), (A); *Shughni* 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, -imae (H), (H); *Sogdian* dn (B), (B); *Kumzari* wāh (E), (E); *Middle Persian* s’r (J), (J); *Zebaki* gal (C), (C); *Zazaki* bi,rey-di,eve (D), (D);
- woman:** *Zazaki* ĵinēk (A), (A); *Balochi* ZAL (A), (A); *Yaghnobi* inčak,zan (A), (A); *Ossetic* binojna? (A), (A); *Waziri* SHEZA (C), (C); *Wakhi* AURUT, OJIZ, XUINUN (B), (B); *Pashto* SEDZA (C), (C); *Ishkashmi* štok (C), (C); *Shughni* awrāt, bēwā, ~γinik, kaxōy (A), (A); *Sogdian* ’stryc (C), (C); *Kurdish* jin, afret (A), (A); *Persian* ZAN (A), (A); *Kumzari* zank (A), (A); *Avestan* ĵainiš, strī, nāiri (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’aed (G), (G); *Khwarezmian* δy(n) (F), (F); *Larestani* jangal, jangal-e kaydū

(D), (D); *Shughni* diraxtzōr,jangal (H), (H); *Avestan* razurəm (A), (A); *Parthian* jn (E), (E); *Waziri* ZANGAL (E), (E);

**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* gūNā,pāra- (G), (G); *Pashto* KIRM (A), (A); *Zazaki* bir (C), (C); *Wakhi* PERIC (A), (A);

**year:** *Shughni* sōl (G), (G); *Yazgulyam* sāuza (H), (H); *Larestani* sāl (B), (B); *Old Persian* θard- (B), (B); *Zebaki* sāl (I), (I); *Zazaki* seři (B), (B); *Waziri* KOL (C), (C); *Wakhi* SOL (B), (B); *Bashkardi* sōr/sūr (B), (B); *Ossetic* afaedz (D), (D); *Ormuri* čān,sul (E), (E); *Yaghnobi* sol,yōso (J), (J); *Balochi* SAL (B), (B); *Ishkashmi* sāl (F), (F); *Persian* SAL (B), (B); *Sogdian* srδδ (A), (A); *Bandari* sāl (B), (B); *Pashto* KAL (C), (C); *Avestan* yārə (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* zyrtyḥ (A), (A);

**you:** *Shughni* tama (A), (A); *Waziri* TUS, TOSE (A), (A); *Ossetic* cmax, cmaxtae (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* yūžəm, 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

Avestan	0100?110?????00000??0??1110111
Old Persian	0000?110?????00000??0??1110101
Wakhi	00011100?????00100??0??11111??
Balochi	000100?10100100001101111010110
Kurdish	0000?0?11000000001100??1100110
Zazaki	0100?0?11010000001101101100101
New Persian	0000?0?11000000011101010??0101
Kumzari	1000?0?1100000001110??0??0110
Bashkardi	1000?0?1100000001110??0??0110
Larestani	0000?0?1100000001110??0??0110
Bandari	0000?0?1100000001110??0??0110
Mazandarani	0100?0?10100000011011110??0101
Pashto	011110?0?????00000??0??11101??
Ossetic	0100?100?????00000??0??11101??
Ormuri	010110?0?????10000??0??10011??
Parachi	010110?0?????00000??0??11101??
Ishkashmi	01111100?????00000??0??11101??
Shughni	0100?110?????00000??0??11111??
Yazgulyam	0100?110?????00000??0??11111??
Yaghnobi	0100?100?????00000??0??11111??
Sogdian	0100?110?????00000??0??11111??
Middle Persian	0000?0?0?????01000??0??11101??
Parthian	0100?0?0?????10000??0??1110101
Khwarezmian	0100?1111000000000??1111110101
Bactrian	0110?0?11001000000??0??1110101
Khotanese Saka	000110?0?????01100??0??11001??

# 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  $*\underset{\sim}{i}$ ,  $*\underset{\sim}{u}$  > 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., *\*zárita-* > *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  $\bar{o}$ ; OP *ai*, *a(h)ya*, *iya* > MP  $\bar{e}$
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 *\*hizwāna-* > (dialectal OP *hizāna-*) > MP *izwān/uzwān* ‘tongue’

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

1. MP *-V̄y*, *-V̄w* >  $\bar{V}$  (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)<sup>1</sup>
5. MP *-ig* > NP *-i*, MP *-ug* > NP *-ū*, 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 *\*-č-* >  $\check{z}$
3. PIr *\*d<sub>u</sub>-* > *b-*
4. PIr *\*-b-*, *\*-d-*, *\*-g-* >  $\beta$ ,  $\delta$ ,  $\gamma$

---

<sup>1</sup>This 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 *maig*; Bal *-g* is irregular; see Korn 2005:164)
2. PIr *\*-č-* is retained as *č*
3. PIr *\*d̥u-* > *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 *\*ā̇(h)ia-* > *a*

## Kurdish

Kurdish undergoes the following relevant developments (?):

1. It undergoes the syncope rule
2. PIr *\*u̇i-* > *gy-* (e.g., *gyān* ‘soul’)
3. PIr *\*-č-* > *ž*
4. PIr *\*d̥u-* > *d*
5. PIr *\*-d-*, *\*-g-* > *y* (between identical low vowels),  $\emptyset$
6. PIr *\*-p-*, *\*-t-*, *\*-k-* > *w*, *y*, *g*
7. PIr *\*ā̇(h)ia-* > *e*

## Mazandarani

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

1. It undergoes the syncope rule
2. PIr *\*-č-* > *j*
3. PIr *\*d̥u-* > *b*
4. PIr *\*-d-* >  $\emptyset$

5. PIr \*-p-, \*-t-, \*-k- > w (between identical low vowels), y (between identical low vowels), g
6. PIr \* $\check{a}(h)\underset{\check{}}{i}a-$  >  $\bar{e}$
7. PIr \*-a- is lengthened to  $\bar{a}$  in some contexts

### Zazaki

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

1. It undergoes the syncope rule
2. PIr \* $\underset{\check{}}{u}\underset{\check{}}{i}-$  > g- (e.g., *gān* ‘soul’)
3. PIr \*-č- > ž
4. PIr \* $\underset{\check{}}{d}\underset{\check{}}{u}-$  > 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 \* $\check{a}(h)\underset{\check{}}{i}a-$  >  $\bar{a}$
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- > ž
2. PIr \* $\underset{\check{}}{d}\underset{\check{}}{u}-$  > b
3. PIr \*-b-, \*-d-, \*-g- > v, y, g
4. PIr \*-p-, \*-t-, \*-k- > v, y, g
5. PIr \* $\check{a}(h)\underset{\check{}}{i}a-$  >  $\bar{e}$

## Yazdi

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

1. It undergoes the syncope rule
2. PIr  $*\check{c}- > j$
3. PIr  $*d\check{u}- > b$
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  $*\check{a}(h)\check{i}a- > \bar{e}$
9. PIr  $*a- > a$  (often written  $\check{a}$  [æ])

## Awromani

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

1. It undergoes the syncope rule
2. PIr  $*\check{u}\check{i}- > gy-$  (e.g.,  $gy\bar{a}n$  ‘soul’)
3. PIr  $*\check{c}- > j$
4. PIr  $*d\check{u}- > b$
5. PIr  $*b-, *d-, *g- > w$  (between identical low vowels),  $y$  (between identical low vowels),  $\gamma$
6. PIr  $*p-, *t-, *k- > w, \delta, \gamma$
7. PIr  $*p-, *t-, *k- > b, t, x$
8. PIr  $*\check{a}(h)\check{i}a- > \bar{e}$
9. PIr  $*a- > a$  (often written  $\check{a}$  [æ])

Table 6.1: Distance measures between expected and observed functional items

Lang	Fnc	Obs	Exp	QC	LD	RLD	LLD
MP	object	rAy	rAy	0	0	0	0
MP	comp	kE	kE	0	0	0	0
MP	something	Ec	Ec	0	0	0	0
MP	here	En gyAg	hEn gyAg	1	1	0.0833333333	0.6931471806
MP	from	az	haz	1	1	0.25	0.6931471806
MP	if	agar	hagar	1	1	0.1428571429	0.6931471806
MP	'with' ba	pad	pad	0	0	0	0
MP	other	dUdIgar	dUdIgar	0	0	0	0
MP	imperfective	hamEv	hamEv	0	0	0	0
MP	adversative	mA agar	mA hagar	1	1	0.1	0.6931471806
MP	one	yak	Ek	2	2	0.3333333333	1.0986122887
MP	into	andar	andar	0	0	0	0
MP	every	harv	harv	0	0	0	0
Pth	object	rAD	rAD	0	0	0	0
Pth	comp	kE	kE	0	0	0	0
Pth	something	EviZ	EviZ	0	0	0	0
Pth	from	aZ	haZ	1	1	0.25	0.6931471806
Pth	if	ag	hagar	2	3	0.4285714286	1.3862943611
Pth	'with' ba	pad	pad	0	0	0	0
Pth	other	bidIgar	bidIgar	1	0	0	0
Pth	imperfective	hamEv	hamEv	0	0	0	0
Pth	one	Ev	Ev	0	0	0	0
Pth	into	andar	andar	0	0	0	0
Pth	every	harv	harv	0	0	0	0
NP	object	rA	rA	0	0	0	0
NP	comp	ki	kI	1	1	0.2	0.6931471806
NP	something	hIc	Ic	1	1	0.1666666667	0.6931471806
NP	here	injA	hInjA	2	2	0.1666666667	1.0986122887
NP	from	az	haz	1	1	0.25	0.6931471806
NP	if	agar	hagar	1	1	0.1428571429	0.6931471806
NP	'with' ba	ba	pad	2	2	0.5	1.0986122887
NP	other	digar	dUdIgar	1	3	0.2307692308	1.3862943611
NP	imperfective	mE	hamE	1	2	0.25	1.0986122887
NP	adversative	magar	mA hagar	2	3	0.3	1.3862943611
NP	one	yak	Ik	2	2	0.3333333333	1.0986122887
NP	into	dar	andar	1	2	0.4	1.0986122887
NP	every	har	har	0	0	0	0
Bal	object	rA	rAd	1	1	0.25	0.6931471806
Bal	comp	ki	kA	2	1	0.2	0.6931471806
Bal	something	hIc	Evac	4	3	0.5	1.3862943611
Bal	from	ac	hac	1	1	0.25	0.6931471806
Bal	if	agar	hakar	2	2	0.2857142857	1.0986122887
Bal	'with' ba	pa	pat	1	1	0.25	0.6931471806
Bal	other	digar	pitIkar	3	4	0.3076923077	1.6094379124
Bal	adversative	magar	mA hakar	3	4	0.4	1.6094379124
Bal	one	yak	Evak	4	2	0.3333333333	1.0986122887
Kd	comp	ka	ka	0	0	0	0
Kd	something	hIc	Ic	1	1	0.1666666667	0.6931471806

Kd	here	injA	hIngyA	2	4	0.3333333333	1.6094379124
Kd	from	Z	haZ	1	2	0.5	1.0986122887
Kd	if	agar	hagar	1	1	0.1428571429	0.6931471806
Kd	'with' ba	ba	pay	2	2	0.5	1.0986122887
Kd	adversative	magar	mA hagar	2	3	0.3	1.3862943611
Kd	one	yak	Ek	2	2	0.3333333333	1.0986122887
Kd	every	har	har	0	0	0	0
Maz	object	ra	ra	0	0	0	0
Maz	comp	ke	kE	1	1	0.2	0.6931471806
Maz	something	hacci	acci	1	1	0.1666666667	0.6931471806
Maz	here	InjA	hInjA	1	1	0.0833333333	0.6931471806
Maz	from	jA	haj	1	3	0.75	1.3862943611
Maz	if	Agar	hAgar	1	1	0.1428571429	0.6931471806
Maz	'with' ba	ba	pay	2	2	0.5	1.0986122887
Maz	adversative	mAgar	mA hagar	2	3	0.3	1.3862943611
Maz	one	yak	atta	3	4	0.6666666667	1.6094379124
Maz	into	dar	andar	1	2	0.4	1.0986122887
Maz	every	har	har	0	0	0	0
Zaz	object	rA	rA	0	0	0	0
Zaz	comp	ki	kA	2	1	0.2	0.6931471806
Zaz	here	InjA	hIngA	3	2	0.1666666667	1.0986122887
Zaz	if	eger	heger	1	1	0.1428571429	0.6931471806
Zaz	'with' ba	bi	pa	2	2	0.5	1.0986122887
Zaz	one	Zev	Zev	0	0	0	0
Zaz	every	her	her	0	0	0	0
Sang	object	re	re	0	0	0	0
Sang	comp	ku	kU	1	1	0.2	0.6931471806
Sang	something	havcu	avcu	1	1	0.1666666667	0.6931471806
Sang	here	anjU	hanjU	1	1	0.0833333333	0.6931471806
Sang	from	az	haZ	1	2	0.5	1.0986122887
Sang	if	aga	haga	1	1	0.1428571429	0.6931471806
Sang	'with' ba	ba	pe	2	2	0.5	1.0986122887
Sang	other	dIgar	bidIgar	1	2	0.1538461538	1.0986122887
Sang	imperfective	mI	hamI	1	2	0.25	1.0986122887
Sang	adversative	maga	mA hagar	2	4	0.4	1.6094379124
Sang	one	yak	Evk	3	2	0.3333333333	1.0986122887
Sang	into	de	ande	1	2	0.4	1.0986122887
Yazdi	object	ra	rad	1	1	0.25	0.6931471806
Yazdi	comp	gI	xI	1	1	0.2	0.6931471806
Yazdi	something	heS	eS	1	1	0.1666666667	0.6931471806
Yazdi	from	az	haj	2	2	0.5	1.0986122887
Yazdi	if	agar	hagar	1	1	0.1428571429	0.6931471806
Yazdi	'with' ba	ba	bat	1	1	0.25	0.6931471806
Yazdi	imperfective	e	hame	2	3	0.375	1.3862943611
Yazdi	adversative	magar	mA hagar	2	3	0.3	1.3862943611
Yazdi	one	yak	Ek	2	2	0.3333333333	1.0986122887
Yazdi	into	dar	edar	1	1	0.2	0.6931471806
Yazdi	every	har	har	0	0	0	0
Awromani	object	ra	ra	0	0	0	0
Awromani	comp	ka	kE	2	1	0.2	0.6931471806
Awromani	something	hIc	Ic	1	1	0.1666666667	0.6931471806

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

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

NP	nAma	name	nAm	1	0.25	0.6931472
NP	JAn	know	dAn	2	0.6666667	1.0986123
NP	JAmAtR	mother-in-law	dAmAd	3	0.5	1.3862944
NP	Traya	three	se	5	1	1.7917595
NP	panca	five	panj	2	0.4	1.0986123
NP	hafta	seven	haft	1	0.2	0.6931472
NP	aSta	eight	haSt	2	0.5	1.0986123
Bal	paca	cook	pac	1	0.25	0.6931472
Bal	kapautaka	blue/gray/pigeon	kapot	5	0.5555556	1.7917595
Bal	taca	run	tac	1	0.25	0.6931472
Bal	bUta	be	bUt	1	0.25	0.6931472
Bal	huSka	dry	huSk	1	0.2	0.6931472
Bal	brAtar	brother	brAt	2	0.3333333	1.0986123
Bal	xumba	pot	kumb	2	0.4	1.0986123
Bal	dauca	sew	dOc	3	0.6	1.3862944
Bal	dantan	tooth	dantAn	2	0.3333333	1.0986123
Bal	xSaudaya	hunger	sOd	7	0.875	2.0794415
Bal	pAda	foot	pAd	1	0.25	0.6931472
Bal	niguSa	listen	nigOS	2	0.3333333	1.0986123
Bal	kaufa	mountain	kOpag	5	1	1.7917595
Bal	kafa	fall	kap	2	0.5	1.0986123
Bal	paTana	bridge	patan	2	0.3333333	1.0986123
Bal	gUTa	excrement	gUt	2	0.5	1.0986123
Bal	maeTa	stay/live	mEta	4	0.8	1.6094379
Bal	xAda	bite/break	kAd	2	0.5	1.0986123
Bal	xAnya	well	kAnIg	4	0.8	1.6094379
Bal	xara	donkey	kar	2	0.5	1.0986123
Bal	cara	move	carr	2	0.5	1.0986123
Bal	caSman	eye	camm	3	0.5	1.3862944
Bal	cina	choose	cin	1	0.25	0.6931472
Bal	wicin	select	gicin	2	0.4	1.0986123
Bal	rauca	light/day	roc	3	0.6	1.3862944
Bal	jana	offspring	jan	1	0.25	0.6931472
Bal	jani	woman	jan	1	0.25	0.6931472
Bal	buja	open	boj	2	0.5	1.0986123
Bal	ACuna	iron	Asin	3	0.6	1.3862944
Bal	sauca	needle	sOc	3	0.6	1.3862944
Bal	paCu	cow	pas	2	0.5	1.0986123
Bal	wahuni	blood	gUn	5	0.8333333	1.7917595
Bal	Jasta	hand	dast	2	0.4	1.0986123
Bal	mAh	moon	mAh	1	0.3333333	0.6931472
Bal	aCanga	iron	sang	3	0.5	1.3862944
Bal	wAta	wind	gwAt	2	0.5	1.0986123
Bal	zRd	heart	zird	3	1	1.3862944
Bal	waic	select	gEc	4	1	1.6094379
Bal	dUta	smoke	dUt	1	0.25	0.6931472
Bal	Ayaka	egg	Ayak	1	0.2	0.6931472
Bal	nAma	name	nav	3	0.75	1.3862944
Bal	JAn	know	zAn	2	0.6666667	1.0986123
Bal	JAmAtR	mother-in-law	zAmAt	2	0.3333333	1.0986123
Bal	Traya	three	say	3	0.6	1.3862944

Bal	panca	five	panc	1	0.2	0.6931472
Bal	hafta	seven	apt	3	0.6	1.3862944
Bal	aSta	eight	aSt	1	0.25	0.6931472
Kd	rauca	light/day	rOZ	4	0.8	1.6094379
Kd	caSman	eye	caw	4	0.6666667	1.6094379
Kd	kaufa	mountain	kOf	3	0.6	1.3862944
Kd	vahuni	blood	xwen	5	0.8333333	1.7917595
Kd	panca	five	penj	3	0.6	1.3862944
Kd	Apa	water	Av	2	0.6666667	1.0986123
Kd	nAma	name	nAv	2	0.5	1.0986123
Kd	dantan	tooth	dan	3	0.5	1.3862944
Kd	vyAna	soul	gyAn	2	0.4	1.0986123
Kd	vAta	wind	bA	3	0.75	1.3862944
Kd	visati	twenty	bIst	4	0.6666667	1.6094379
Maz	rauca	light/day	rUj	4	0.8	1.6094379
Maz	caSman	eye	caS	3	0.5	1.3862944
Maz	kaufa	mountain	kOh	4	0.8	1.6094379
Maz	vahuni	blood	xUn	5	0.8333333	1.7917595
Maz	panca	five	panj	2	0.4	1.0986123
Maz	gauSa	ear	gUS	3	0.6	1.3862944
Maz	Apa	water	U	3	1	1.3862944
Maz	nAma	name	nAm	1	0.25	0.6931472
Maz	dantan	tooth	dAnUn	4	0.6666667	1.6094379
Maz	vyAna	soul	jAn	4	0.8	1.6094379
Maz	vAta	wind	vAd	2	0.5	1.0986123
Maz	visati	twenty	vIst	3	0.5	1.3862944
Maz	pitar	father	per	4	0.8	1.6094379
Zaz	rauca	light/day	rOZ	4	0.8	1.6094379
Zaz	caSman	eye	Cim	5	0.8333333	1.7917595
Zaz	kaufa	mountain	ko	4	0.8	1.6094379
Zaz	vahuni	blood	gUnI	6	1	1.9459101
Zaz	panca	five	panj	2	0.4	1.0986123
Zaz	gauSa	ear	vIst	5	1	1.7917595
Zaz	Apa	water	Aw	2	0.6666667	1.0986123
Zaz	nAma	name	nAm	1	0.25	0.6931472
Zaz	dantan	tooth	dindAn	4	0.6666667	1.6094379
Zaz	vyAna	soul	gAn	3	0.6	1.3862944
Zaz	vAta	wind	vA	2	0.5	1.0986123
Zaz	visati	twenty	vIst	3	0.5	1.3862944
Zaz	pitar	father	pI	4	0.8	1.6094379
Sang	rauca	light/day	rUZ	4	0.8	1.6094379
Sang	caSman	eye	caS	3	0.5	1.3862944
Sang	kaufa	mountain	kUh	4	0.8	1.6094379
Sang	vahuni	blood	xUn	5	0.8333333	1.7917595
Sang	panca	five	panj	2	0.4	1.0986123
Sang	gauSa	ear	gUS	3	0.6	1.3862944
Sang	Apa	water	Ov	3	1	1.3862944
Sang	nAma	name	nUm	2	0.5	1.0986123
Sang	dantan	tooth	dannUn	3	0.5	1.3862944
Sang	vyAna	soul	Jun	4	0.8	1.6094379
Sang	vAta	wind	ve	3	0.75	1.3862944



Sang	visati	twenty	vIst	3	0.5	1.3862944
Yzd	rauca	light/day	rUj	4	0.8	1.6094379
Yzd	paTana	bridge	pAn	4	0.6666667	1.6094379
Yzd	caSman	eye	cam	3	0.5	1.3862944
Yzd	vahuni	blood	xin	5	0.8333333	1.7917595
Yzd	panca	five	panj	2	0.4	1.0986123
Yzd	gauSa	ear	gUS	3	0.6	1.3862944
Yzd	Apa	water	wuw	4	1.3333333	1.6094379
Yzd	dantan	tooth	dedun	5	0.8333333	1.7917595
Yzd	vyAna	soul	Jun	4	0.8	1.6094379
Yzd	vAta	wind	vUd	3	0.75	1.3862944
Yzd	visati	twenty	vist	2	0.3333333	1.0986123
Yzd	pitar	father	bder	5	1	1.7917595
Awr	rauca	light/day	rO	4	0.8	1.6094379
Awr	caSman	eye	cam	3	0.5	1.3862944
Awr	vahuni	blood	vun	3	0.5	1.3862944
Awr	panca	five	panj	2	0.4	1.0986123
Awr	gauSa	ear	gOs	4	0.8	1.6094379
Awr	Apa	water	Awr	3	1	1.3862944
Awr	nAma	name	nAm	1	0.25	0.6931472
Awr	dantan	tooth	daDAn	4	0.6666667	1.6094379
Awr	vyAna	soul	gyAn	2	0.4	1.0986123
Awr	visati	twenty	bist	3	0.5	1.3862944

# 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 = Naṣrī Ašrafī 2002; Miller = Miller 1892; Rossi = Rossi 1998.

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

\*u(a)rda- ‘flower’: *New Persian* gul; *Larestani* gol; *Balochi* gul; *Sangesari* gol; *Yazdi* gulūw rosewater; *Zazaki* gul; *Sivandi* gol; *Central Dialects* N ; *S Tati* vela p. 68; *Kurdish* gul; *Mazandarani* gəl FVT

\*špržan- ‘spleen’: *New Persian* supurz; *Sangesari* espol; *Yazdi* sebul, svarz; *Sivandi* espel; *S Tati* supurz Miller; *Kurdish* sipil

\*brž- ‘high’: *New Persian* buland,burz; *Balochi* burz; *Sangesari* belënd; *Yazdi* belend; *Zazaki* berz; *Sivandi* boland; *Central Dialects* N bilénd; *S Tati* bulund Miller; *Kurdish* barz,bilind; *Mazandarani* belan

\*žar(i)t- ‘yellow/gold’: *New Persian* zard,dal; *Larestani* zard; *Balochi* zard; *Sangesari* zar; *Yazdi* zart; *Sivandi* zard; *Central Dialects* Q zärd

\*marž- ‘wipe’: *New Persian* māḷīdan; *Larestani* māleda (rub); *Sivandi* māl-; *Central Dialects* Kh māl-; *Kurdish* māl

\*harž- ‘release’: *New Persian* hālīdan; *Zazaki* erzen; *Sivandi* äl; *Central Dialects* Q hel-; *S Tati* harz; *Kurdish* pāl

\*šata- ‘hundred’: *New Persian* sad; *Balochi* sad; *Sangesari* sey; *Yazdi* sad; *Sivandi* sad; *Kurdish* säd; *Mazandarani* sad

\*laupāša- ‘fox’: *New Persian* rubāh; *Balochi* rūbā; *Sangesari* reva; *Yazdi* ruwās; *Sivandi* rūbā; *Central Dialects* rūbā; *S Tati* luās; *Kurdish* Rewi

- \**daša-* ‘ten’: *New Persian* dah; *Larestani* da; *Balochi* da; *Sangesari* das; *Yazdi* dah; *Sivandi* da; *S Tati* da(h) p. 144; *Kurdish* də; *Mazandarani* da
- \**kašu-*/*\*kašiah-* ‘small/er’: *New Persian* kih; *Sangesari* kas; *Yazdi* kas
- \**uināša-* ‘sin’: *New Persian* gunāh; *Balochi* gunā; *Yazdi* gunū; *Sivandi* gonā; *Central Dialects* Q gunāh
- \**ašu(a)na-* ‘iron’: *New Persian* āhen; *Larestani* ā’en; *Balochi* ā(h)in; *Sangesari* @hən; *Yazdi* wuhen; *Zazaki* hesinīn of iron; *Sivandi* āhān; *Kurdish* hesin; *Mazandarani* ā’en FVT
- \**šuiš-* ‘louse/flea’: *New Persian* šepeš; *Larestani* eš; *Sangesari* εspaz; *Yazdi* šeweš; *S Tati* espeja p. 69; *Mazandarani* espij
- \**ašua-* ‘horse’: *New Persian* savār,asb; *Balochi* asp; *Sangesari* asm; *Yazdi* asb; *Sivandi* usūr; *Central Dialects* Q äs; *S Tati* asb p. 127, suār horse men, p. 39; *Kurdish* əsp; *Mazandarani* asb
- \**šuaka-* ‘dog’: *New Persian* sag; *Sangesari* εspe; *Yazdi* sve; *Sivandi* espe; *S Tati* esbi p. 80; *Kurdish* səg; *Mazandarani* sag
- \**kašīapa-* tortoise/turtle: *New Persian* kašaf; *Balochi* kasīb; *Sangesari* l@kpa\$; *Sivandi* kalapošt; *S Tati* kasuya, kasawa p. 69; *Kurdish* küsel Behd
- \**šīāua-* ‘black’: *New Persian* siyāh; *Balochi* siyā; *Sangesari* so; *Yazdi* siyū; *Zazaki* siyā; *Sivandi* siya; *Central Dialects* N sīyāh; *S Tati* siā p. 262; *Mazandarani* siv
- \**māšīa-* ‘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* māī FVT
- \**urka-* wolf: *New Persian* gurg; *Balochi* gurk; *Sangesari* vark; *Yazdi* gorg; *Central Dialects* Semn vārg, N gurg; *S Tati* varg; *Kurdish* gurg; *Mazandarani* varg
- \**uahu-* ‘good’: *New Persian* beh(tar); *Sangesari* veytar better; *Yazdi* vatār better; *Zazaki* weš; *Sivandi* va(htar); *Central Dialects* M vêtār
- \**uata-* ‘bad’: *New Persian* bad; *Balochi* bad; *Sangesari* bad; *Yazdi* vad; *Sivandi* bad; *Central Dialects* bād; *Kurdish* bəd
- \**uāta-* ‘wind’: *New Persian* bād; *Larestani* bād; *Balochi* gwāt; *Sangesari* ve; *Yazdi* wūd; *Zazaki* vā; *Sivandi* voy; *Central Dialects* vā; *S Tati* vār (CLI); *Kurdish* ba; *Mazandarani* va:
- \**uašīah-* (?) ‘more’: *New Persian* beš; *Balochi* geš; *Sangesari* ve\$ter; *Yazdi* veštar; *Sivandi* vīštar; *Central Dialects* N vēshtir

\**ur̥tka*- ‘kidney’: *New Persian* gurda; *Larestani* gordakü; *Balochi* guTTig (Rossi); *Sangesari* vakku; *Sivandi* velk; *Central Dialects* N gurdä; *Mazandarani* ga:rda

\**ui(n)šati*- ‘twenty’: *New Persian* bīst; *Larestani* biss; *Balochi* gīst, bīst; *Sangesari* vist; *Yazdi* bīst; *Sivandi* vīs; *S Tati* vist; *Kurdish* bist; *Mazandarani* bist

\**uar(ša)*- ‘rain’: *New Persian* bārān; *Larestani* baru; *Balochi* gwārān; *Sangesari* v@r rains; *Yazdi* wūrun; *Zazaki* vārān; *Sivandi* vārān; *Central Dialects* Kh vārūn; *S Tati* vāriš (CLI); *Kurdish* baran; *Mazandarani* va:reš

\**uafra*- ‘ice/snow’: *New Persian* barf; *Larestani* vafr, barf; *Balochi* barp; *Sangesari* varf~vafr; *Yazdi* wapr; *Zazaki* vewr; *Sivandi* varf; *Central Dialects* Q varf; *S Tati* Tk vara; *Kurdish* bəfr; *Mazandarani* varf

Pair	greater_circle(m)	least_cost(m)	minute	all	*r/rt/ž/d	*ś(w/y)	*w-
NP-Lar	932389.85	2540235.94	15420	4.623611	2.738613	2.473388	2.872281
NP-Bal	1576679.28	4145316.02	24240	5.110543	2.645751	3.162278	2.924988
NP-Kd	670979.10	1758158.98	11700	3.995095	2.148345	3.372684	0.000000
NP-Zaz	1098966.23	2619390.27	17760	6.836435	2.631174	4.376973	4.595650
NP-Tat	161104.44	444885.17	2340	5.964639	1.167748	3.824265	4.449719
NP-Yzd	501477.51	1606831.67	7620	5.702206	2.645751	3.007238	4.062019
NP-Siv	638734.86	1654773.91	9840	4.966555	2.000000	2.549510	3.741657
NP-Maz	177039.08	1040002.91	4860	5.412178	1.581139	3.224903	3.872983
NP-Sng	174513.48	332857.94	2820	5.876507	2.148345	3.509386	4.204589
NP-Xun	177862.35	374167.46	2580	4.580947	1.860521	1.927248	3.605551
Lar-Bal	1264982.69	4164102.61	22200	4.966555	3.354102	2.142017	3.146427
Lar-Kd	1358365.75	3226053.74	21540	5.844129	3.535534	4.000000	2.872281
Lar-Zaz	1809578.70	4152302.69	27360	6.875884	3.354102	4.000000	4.449719
Lar-Tat	1031479.07	2351052.19	16080	5.520017	2.236068	2.174229	4.405878
Lar-Yzd	469777.02	1269392.89	7800	5.608030	1.936492	3.029269	4.449719
Lar-Siv	301345.90	893874.32	5760	5.408864	1.936492	2.633122	4.260282
Lar-Maz	995827.23	3505585.47	16680	5.089204	2.236068	2.633122	3.727564
Lar-Sng	898785.81	2798440.50	15060	5.866737	1.936492	3.224903	4.449719
Lar-Xun	881892.12	2825032.44	14820	5.373830	2.236068	2.943920	3.853570
Bal-Kd	2238271.62	5767923.90	35280	4.507075	1.519109	3.372684	2.738613
Bal-Zaz	2675627.83	6629028.18	41340	6.373967	2.148345	3.879772	4.751731
Bal-Tat	1737676.35	4454650.09	26520	6.213590	2.860388	2.549510	4.931104
Bal-Yzd	1219496.69	4069184.14	20160	5.770615	2.828427	3.007238	4.062019
Bal-Siv	1356593.91	4107039.21	24420	5.502066	2.236068	3.122499	3.986087
Bal-Maz	1475222.42	4537322.90	22380	5.469451	3.162278	2.792848	3.760699
Bal-Sng	1413912.60	3859641.81	21840	6.029169	2.401922	3.094987	4.595650
Bal-Xun	1405255.74	3909125.92	21660	5.395809	2.401922	2.943920	3.829708
Kd-Zaz	458267.14	1722073.54	6420	6.480741	2.611165	4.163332	4.405878
Kd-Tat	512596.02	1421509.86	9540	6.523036	2.611165	3.605551	4.751731
Kd-Yzd	1067823.51	2600751.56	16560	6.426689	2.631174	4.472136	3.940737
Kd-Siv	1071971.99	2355354.75	17340	5.893615	1.860521	4.309458	3.872983
Kd-Maz	812892.10	2662610.79	15060	6.301374	2.738613	4.309458	3.760699
Kd-Sng	843936.47	1955465.82	14520	6.353238	2.022600	4.861724	4.062019
Kd-Xun	848622.87	1996775.34	14280	5.229515	1.860521	3.437758	3.674235

Zaz-Tat	937952.08	2308467.45	15660	4.524167	2.860388	2.828427	2.074692
Zaz-Yzd	1524248.70	3527000.50	22620	4.966555	2.841993	3.308681	2.297825
Zaz-Siv	1527747.29	3281603.69	23400	5.581926	3.038218	3.699218	2.814249
Zaz-Maz	1215943.90	3523715.08	21120	4.412900	3.162278	2.633122	1.970369
Zaz-Sng	1264737.69	2816570.11	20580	4.645787	2.860388	3.271984	1.284523

# Bibliography

- Abaev, V. I. (1989). *Istoriko-étimologičeskij slovar' osetinskogo jazyka*, Volume 1-4. Moscow: Nauka.
- Adiego Lajara, I.-X. (2000). Sobre el relativo del antiguo persa. In B. Forssman and R. Plath (Eds.), *Indoarisch, Iranisch und die Indogermanistik: Arbeitstagung der Indogermanischen Gesellschaft vom 2. bis 5. Oktober 1997 in Erlangen*, pp. 1–14. Wiesbaden: Ludwig Reichert Verlag.
- Aikhenvald, A. Y. (2006). Areal diffusion, genetic inheritance, and problems of subgrouping: A North Arawak case study. In A. Y. Aikhenvald and R. M. W. Dixon (Eds.), *Areal Diffusion in Genetic Inheritance*, pp. 167–194. Oxford: Oxford University Press.
- Andreev, M. and E. Peščereva (1957). *Jagnobskije teksty*. Moscow: Izdatel'stvo Akademij Nauk SSSR.
- Asatrian, G. (2012). Marginal remarks on the history of some Persian words. *Iran and the Caucasus* 16, 105–116.
- Azami, C. A. and G. Windfuhr (1972). *A Dictionary of Sangesari with a Grammatical Outline*. Tehran: Franklin Book Programs.
- Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Babel, M., A. Garrett, M. Houser, and M. Toosarvandani (2013). Descent and diffusion in language diversification: A study of Western Numic dialectology. *International Journal of American Linguistics* 79(4), 445–89.
- Bailey, H. W. (1933). Western Iranian dialects. *Transactions of the Philological Society*, 46–64.
- Bailey, H. W. (1936). Yazdi. *Bulletin of the School of Oriental and African Studies* 8(2-3), 335–361.
- Bailey, H. W. (1945). Asica. *Transactions of the Philological Society*, 1–38.

- Bailey, H. W. (1967). *Prolexis to the Book of Zambasta*, Volume 6 of *Khotanese Texts*. Cambridge: Cambridge University Press.
- Bailey, H. W. (1979). *Dictionary of Khotan Saka*. Cambridge: Cambridge University Press.
- Bakker, P., A. Daval-Markussen, I. Plag, and M. Perkvall (2011). Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26(1), 5–42.
- Barker, M. A. (1969). *A Course in Baluchi*. Montreal: Institute of Islamic Studies, McGill University.
- Bartholomae, C. (1895). Vorgeschichte der iranischen Sprachen. In W. Geiger and E. Kuhn (Eds.), *Grundriss der iranischen Philologie*, Volume 1, Chapter 1, pp. 1–151. Strassburg [Strasbourg]: Karl J. Trübner.
- Bates, D., M. Maechler, B. Bolker, and S. Walker (2013). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.0-5.
- Beekes, R. S. P. (1988). *A Grammar of Gatha-Avestan*. Leiden: Brill.
- Beekes, R. S. P. (1997). Historical phonology of Iranian. *Journal of Indo-European Studies* 25(1-2), 1–26.
- Benedictsen, Å. M. and A. Christensen (1921). *Les dialectes d'Awromān et de Pāwā*, Volume 6 of *Historisk-filosofiske Meddelelser*. Copenhagen: Det Kgl. Danske Videnskabernes Selskab.
- Benveniste, E. (1935). *Les infinitifs avestiques*. Paris: Adrien Maisonneuve.
- Benveniste, E. (1966). *Titres et Noms Propres en Iranien Ancien*. Paris: Klincksieck.
- Bergsland, K. and H. Vogt (1962). On the validity of glottochronology. *Current Anthropology* 3(2), 115–153.
- Bielmeier, R. (1989). Yaghnobi. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 480–488. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Blažek, V. (2013). On classification of Middle Iranian languages (preliminary report). *Linguistica Brunensia* 61(1-2), 49–74.
- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart and Winston.
- Borjjan, H. (2008). Tabarica II: some Mazandarani verbs. *Iran and the Caucasus* 12(1), 73–81.
- Borjjan, H. (2009). Median succumbs to Persian after three millennia of coexistence: Language shift in the Central Iranian Plateau. *Journal of Persianate Studies* 2, 62–87.
- Borjjan, H. and M. Borjjan (2008). Ethno-linguistic materials from South Caspian rainforests. *Iranian Studies* 41(3), 365–401.

- Bouckaert, R., P. Lemey, M. Dunn, S. J. Greenhill, A. V. Alekseyenko, A. J. Drummond, R. D. Gray, M. A. Suchard, and Q. D. Atkinson (2012). Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097), 957–960.
- Boyce, M. (1964). The use of relative particles in Middle Iranian. In *Indo-Iranica: mélanges présentés à Georg Morgenstierne, à l’occasion de son soixante-dixième anniversaire*, pp. 28–47. Wiesbaden: Harrassowitz.
- Boyce, M. (1977). *A Word-List of Manichaean Middle Persian and Parthian*, Volume 9a of *Acta Iranica*. Leiden: Brill.
- Cantera, A. (2009). On the history of the Middle Persian nominal inflection. In W. Sundermann, A. Hintze, and F. de Blois (Eds.), *Exegisti Monumenta: Festschrift in Honour of Nicholas Sims-Williams*, Volume 17 of *Iranica*, pp. 17–30. Wiesbaden: Harrassowitz.
- Cathcart, C. A. (forthcoming). Degrees of irregular change: phonological reduction and grammatical convergence in West Iranian. *Language Dynamics and Change* 5(2).
- Chang, W., C. Cathcart, D. Hall, and A. Garrett (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European Steppe Hypothesis. *Language* 91(1), 194–244.
- Cheung, J. (2002). *Studies in the Historical Development of the Ossetic Vocalism*, Volume 21 of *Beiträge zur Iranistik*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Cheung, J. (2007). *Etymological dictionary of the Iranian verb*. Leiden: Brill.
- Clackson, J. (2007). *Indo-European Linguistics*. Cambridge: Cambridge University Press.
- Collett, N. (1909). *A Grammar, Phrase Book and Vocabulary of Baluchi: (as spoken in the Sultanate of Oman)* (2nd ed.). Camberley: s.n.
- Cowgill, W. (1970). Italic and Celtic superlatives and the dialects of Indo-European. In G. Cardona, H. M. Hoenigswald, and A. Senn (Eds.), *Indo-European and Indo-Europeans: Papers presented at the Third Indo-European Conference at the University of Pennsylvania*, pp. 113–154. Philadelphia: University of Pennsylvania Press.
- Darlington, R. B. (1970). Is kurtosis really “peakedness”? *The American Statistician* 24(2), 19–22.
- de Ujfalvy, C. (1896). *Les Aryens au nord et au sud de Hindou-Kouch*. Paris: G. Masson.
- de Vaan, M. (2003). *The Avestan Vowels*. Amsterdam: Rodopi.
- Demiraj, B. (1997). *Albanische Etymologien*. Amsterdam: Rodopi.
- Donohue, M. (2012). Typology and areality. *Language Dynamics and Change* 2, 98–116.



- Donohue, M., T. Denham, and S. Oppenheimer (2012). New methodologies for historical linguistics? Calibrating a lexicon-based methodology for diffusion vs. subgrouping. *Diachronica* 29(4), 505–522.
- Donohue, M., S. Wichmann, and M. Albu (2008). Typology, areality, and diffusion. *Oceanic Linguistics* 47(1), 223–232.
- Dunn, M. (2015). Language phylogenies. In C. Bower and B. Evans (Eds.), *The Routledge Handbook of Historical Linguistics*, pp. 190–211. New York; Oxford: Routledge.
- Dunn, M., S. C. Levinson, E. Lindström, G. Reesink, and A. Terrill (2008). Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language* 84(4), 710–759.
- Dunn, M. and J. Ludewig (2012). IELEX: Indo-european lexical cognacy database.
- Durkin-Meisterernst, D. (2004). *Dictionary of Manichaean Texts III*. Turnhout: Brepols.
- Dyen, I., J. Kruskal, and P. Black (1992). An Indoeuropean classification: a lexicostatistical experiment. *Transactions of the American Philosophical Society* 82(5).
- Édel'man, D. I. (1980). History of the consonant systems of the North-Pamir languages. *Indo-Iranian Journal* 22, 287–310.
- Édel'man, D. I. (1992). Ešče raz ob etapax filiacii arijskoj jazykovoju obščnosti. *Voprosy Jazykoznanija* (3), 44–66.
- Édel'man, D. I. and L. Dodykhudoeva (2009). Shughni. In G. Windfuhr (Ed.), *The Iranian languages*, pp. 787–824. London: Routledge.
- Elfenbein, J. (1963). *A vocabulary of Marw Baluchi*. Naples: Istituto Universitario Orientale di Napoli.
- Elfenbein, J. (1989). Baloči. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 350–362. Wiesbaden: Ludwig Reichert Verlag.
- Emmerick, R. (1989). Khotanese and Tumshuqese. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 204–229. Wiesbaden: Ludwig Reichert Verlag.
- Fay, E. W. (1895, April 18-19). Avestan *hizva* in Sanskrit. In *Proceedings of the American Oriental Society, at Its Meeting in New Haven, Conn.*, Volume 16 of *Journal of the American Oriental Society*, pp. ccxxviii–ccxxix. American Oriental Society.
- François, A. (2011). Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics* 1(2), 175–246.

- François, A. (2015). Trees, waves and linkages: models of language diversification. In C. Bowerman and B. Evans (Eds.), *The Routledge Handbook of Historical Linguistics*, pp. 161–189. Oxford: Routledge.
- Garrett, A. (2006). Convergence in the formation of Indo-European subgroups: Phylogeny and chronology. In P. Forster and C. Renfrew (Eds.), *Phylogenetic methods and the prehistory of languages*, pp. 139–151. Cambridge: McDonald Institute for Archaeological Research.
- Garrett, A. and K. Johnson (2013). Phonetic bias in sound change. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization*, pp. 51–97. Oxford: Oxford University Press.
- Geraghty, P. A. (1983). *The History of the Fijian Languages*. Number 19 in Oceanic Linguistics Special Publication. Honolulu: University of Hawaii Press.
- Gershevitch, I. (1952). Ancient survivals in Ossetic. *Bulletin of the School of Oriental and African Studies* 14 (3), 483–495.
- Gershevitch, I. (1954). *A Grammar of Manichaean Sogdian*. Oxford: Blackwell.
- Gershevitch, I. (1962). Dialect variation in early Persian. *Transactions of the Philological Society* 63, 1–29.
- Gharib, B. (1995). *Sogdian Dictionary: Sogdian-Persian-English*. Tehran: Farhangian Publications.
- Gilliéron, J. and M. Roques (1912). *Etudes de géographie linguistique d'après l'Atlas linguistique de la France*. Paris: H. Champion.
- Gippert, J. (1993). *Iranica Armeno-Iberica. Studien zu den iranischen Lehnwörtern im Armenischen und Georgischen*. Number 2 vols. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Goddard, I. (1975). Algonquian, Wiyot and Yurok: proving a distant genetic relationship. In M. D. Kinkade, K. L. Hale, and O. Werner (Eds.), *Linguistics and anthropology in honor of C.F. Voegelin*, pp. 249–262. Lisse: Peter de Ridder Press.
- Goebel, H. (2006). Recent advances in Salzburg dialectometry. *Literary and Linguistic Computing* 21 (4), 411–435.
- Gray, R. D. and Q. D. Atkinson (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426 (6965), 435–439.
- Grierson, G. A. (1920). *Ishkashmi, Zebaki, and Yazghulami: an account of three Eranian dialects*. London: Royal Asiatic Society.

- Hamed, M. B. and S. Flavier (2007). UNIDIA: A database for deriving diachronic universals. In M. Dufresne, F. Dupuis, and E. Vocaj (Eds.), *Historical Linguistics 2007: Selected papers from the 18th International Conference on Historical Linguistics*, pp. 259–268. Amsterdam: John Benjamins. Database available online at <http://www.diadm.ish-lyon.cnrs.fr/unidia/>.
- Harmatta, J. (1970). *Studies in the History and Language of the Sarmatians*, Volume 13 of *Acta antiquae et archaeologica*. Szeged: Acta Universitatis de Attila József Nominatae.
- Haynie, H. (2012). *Studies in the History and Geography of California Languages*. Ph. D. thesis, University of California, Berkeley.
- Heggarty, P. (2000). Quantifying change over time in phonetics. In C. Renfrew (Ed.), *Time Depth in Historical Linguistics*, pp. 531–562. Cambridge: McDonald Institute for Archaeological Research.
- Heine, B. and T. Kuteva (2009). *The genesis of grammar: a reconstruction*. Oxford: Oxford University Press.
- Heled, J. and R. Bouckaert (2013). Looking for trees in the forest: summary tree from posterior samples. *BMC Evolutionary Biology* 13, 221.
- Henning, W. B. (1958). Mitteliranisch. *Iranistik* 1(4), 20–130.
- Henning, W. B. (1963a). Coriander. *Asia Major* 10(2), 195–199.
- Henning, W. B. (1963b). The Kurdish elm. *Asia Major* 10(1), 68–72.
- Hintze, A. (1998). The migrations of the Indo-Iranians and the Iranian sound change  $n > s$ . In W. Meid (Ed.), *Sprache und Kultur der Indogermanen. Akten der 10. Fachtagung der Indogermanischen Gesellschaft, Innsbruck 22–29.9.1996.*, Volume 93 of *Innsbrucker Beiträge zur Sprachwissenschaft*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Hoenigswald, H. M. (1960). *Language change and linguistic reconstruction*. Chicago: University of Chicago Press.
- Hoenigswald, H. M. (1966). Criteria for the subgrouping of languages. In H. Birnbaum and J. Puhvel (Eds.), *Ancient Indo-European dialects*, pp. 1–12. Berkeley, CA: University of California Press.
- Hoffmann, K. (1958 [1976]a). Altiranisch. In J. Narten (Ed.), *Karl Hoffmann: Aufsätze zur Indoiranistik*, Volume 1, pp. 58–76. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Hoffmann, K. (1976b). Zur altpersischen schrift. In J. Narten (Ed.), *Karl Hoffmann: Aufsätze zur Indoiranistik*, Volume 2, pp. 620–645. Wiesbaden: Ludwig Reichert Verlag.

- Hoffmann, K. and B. Forssman (2004). *Avestische Laut- und Flexionslehre* (2nd ed.), Volume 84 of *Innsbrucker Beiträge zur Sprachwissenschaft*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Horn, P. (1893). *Grundriss der neupersischen Etymologie*. Strassburg [Strasbourg]: Karl J. Trübner.
- Horn, P. (1901). Neupersische Schriftsprache. In W. Geiger and E. Kuhn (Eds.), *Grundriss der iranischen Philologie*, Volume 1, pp. 1–200. Strassburg: Karl J. Trübner.
- Hübschmann, H. (1897). *Armenische Grammatik*. Leipzig: Breitkopf & Härtel.
- Humbach, H. (1989). Choresmian. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 193–203. Wiesbaden: Ludwig Reichert Verlag.
- Humbach, H. and P. Ichaporia (1998). *Zamyād Yasht: Yasht 19 of the Younger Avesta: text, translation, commentary*. Wiesbaden: Harrassowitz.
- Huson, D. H. and D. Bryant (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2), 254–267.
- Inslar, S. (1971). Vedic *mith*. *Transactions of the Philological Society*, 163–174.
- Ivanow, W. (1940). *The Gabri dialect spoken by the Zoroastrians of Persia*, Volume 16 of *Rivista degli Studi Orientali*. Rome: Scuola Orientale nella R. Università di Roma.
- Jamison, S. (2009). Sociolinguistic remarks on the Indo-Iranian *\*-ka-* suffix: A marker of colloquial register. *Indo-Iranian Journal* 52(2-3), 311–329.
- Joseph, B. D. (2004). Rescuing traditional (historical) linguistics from grammaticalization theory. In O. Fischer, M. Norde, and H. Perridon (Eds.), *Up and Down the Cline: The Nature of Grammaticalization*, pp. 45–71. Amsterdam: John Benjamins.
- Joseph, B. D. (2012). A variationist solution to apparent copying across related languages. In L. Johanson and M. Robbeets (Eds.), *Copies versus Cognates in Bound Morphology*, Volume 2 of *Brill's Studies in Language, Cognition and Culture*, pp. 151–167. Leiden: Brill.
- Josephson, J. (2013). The historical background of the transfer of a Kurdish bound morpheme to Neo-Aramaic. In L. Johanson and M. Robbeets (Eds.), *Copies versus Cognates in Bound Morphology*, pp. 355–370. Leiden: Brill.
- Kalyan, S. and A. François (forthcoming). Freeing the comparative method from the tree model: a framework for historical glottometry. In R. Kikusawa and L. Reid (Eds.), *Let's talk about trees: tackling problems in representing phylogenetic relationships among languages*. Osaka: National Museum of Ethnology.
- Kamioka, K. and M. Yamada (1979). *Lārestani Studies*, Volume 1. Tokyo: Institute for the Study of Cultures of Asia and Africa.

- Kellens, J. and E. Pirart (1988-1991). *Les textes vieil-avestiques*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Kent, R. (1943). Studies in Old Persian morphology. *Language* 19(3), 221–229.
- Kent, R. (1951). *Old Persian*, Volume 33 of *American Oriental Series*. New Haven: American Oriental Society.
- Kessler, B. (1995). Computational dialectology in Irish Gaelic. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*, Dublin, pp. 60–67. EACL.
- Kieffer, C. (1989). Le parāčī, l'ōrmuṛī. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 445–455. Wiesbaden: Ludwig Reichert Verlag.
- Kim, R. (2003). On the historical phonology of Ossetic: The origin of the oblique case suffix. *Journal of the American Oriental Society* 123(1), 43–72.
- Klingenschmitt, G. (2000). Mittelpersisch. In B. Forssman and R. Plath (Eds.), *Indoarisch, Iranisch und die Indogermanistik: Arbeitstagung der Indogermanischen Gesellschaft vom 2. bis 5. Oktober 1997 in Erlangen*, pp. 191–229. Wiesbaden: Ludwig Reichert Verlag.
- Kobayashi, M. (2004). *Historical Phonology of Old Indo-Aryan Consonants*. Tokyo: Fujiwara.
- Korn, A. (2003). Balochi and the concept of North-Western Iranian. In C. Jahani and A. Korn (Eds.), *The Baloch and their neighbours: ethnic and linguistic contacts in Balochistan in historical and modern times*, pp. 49–60. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Korn, A. (2005). *Towards a Historical Grammar of Balochi*. Wiesbaden: Ludwig Reichert Verlag.
- Krahnke, K. (1976). *Linguistic Relationships in Central Iran*. Ph. D. thesis, University of Michigan.
- Kuiper, F. B. J. (1976). Old East Iranian dialects. *Indo-Iranian Journal* 18, 241–253.
- Kümmel, M. (2007). *Konsonantenwandel*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Kümmel, M. (2013 [2012]). The Iranian reflexes of Proto-Iranian \*ns. *Orientalia Suecana* 59, 138–145.
- Lazard, G. (1957). *Grammaire du persain contemporaine*. Paris: Klincksieck.
- Lazard, G. (1963). *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck.
- Lecoq, P. (1979). *Le dialecte de Sivand*, Volume 10 of *Beiträge zur Iranistik*. Wiesbaden: Dr. Ludwig Reichert Verlag.

- Lecoq, P. (1989). Les dialectes caspiens et du nord-ouest. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 296–312. Wiesbaden: Ludwig Reichert Verlag.
- Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard (2009). Bayesian phylogeography finds its roots. *PLOS Computational Biology* 5, e1000520.
- Lentz, W. (1926). Die nordiranischen Elemente in der neupersischen Literatursprache bei Firdosi. *Zeitschrift für Indologie und Iranistik* 4, 251–316.
- Lightfoot, D. (2000). How long was the nineteenth century? (Cuanto tempo durou o século XIX?). *D.E.L.T.A.* 16, 81–98.
- Lipp, R. (2009). *Die indogermanischen und einzelsprachlichen Palatale im Indoiranischen*. Heidelberg: Carl Winter. 2 vols.
- Lommel, H. (1922). Kleine Beiträge zur arischen Sprachkunde. *Kuhns Zeitschrift* 50(3/4), 260–275.
- Lubotsky, A. (2001). The Indo-Iranian substratum. In C. Carpelan, A. Parpola, and P. Koskikallio (Eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations. Papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki 8-10 January 1999*, Helsinki, pp. 301–317.
- Lubotsky, A. (2002). Scythian elements in Old Iranian. In N. Sims-Williams (Ed.), *Indo-Iranian Languages and Peoples*, Volume 116 of *Proceedings of the British Academy*, Oxford, pp. 189–202. Oxford University Press.
- MacKenzie, D. N. (1961). The origins of Kurdish. *Transactions of the Philological Society*, 68–86.
- MacKenzie, D. N. (1971a). *A Concise Pahlavi Dictionary*. London: Oxford University Press.
- MacKenzie, D. N. (1971b). The Khwarezmian glossary—iv. *Bulletin of the School of Oriental and African Studies* 34(3), 521–537.
- MacKenzie, D. N. (2003). The missing link. In L. Paul (Ed.), *Persian Origins: Early Judaeo-Persian and the Emergence of New Persian*, pp. 103–110. Wiesbaden: Harrassowitz.
- Maddison, D. R., D. L. Swofford, and W. P. Maddison (1997). Nexus: an extensible file format for systematic information. *Systematic Biology* 46(4), 590–621.
- Matisoff, J. (2000). On the uselessness of glottochronology for the subgrouping of tibeto-burman. In C. Renfrew, A. McMahon, and L. Trask (Eds.), *Time Depth in Historical Linguistics*, Volume 1, pp. 333–371. Cambridge: McDonald Institute for Archaeological Research.

- Matisoff, J. (2003). *The Handbook of Proto-Tibeto-Burman*. Berkeley, CA: University of California Press.
- Mayrhofer, M. (1984). Lassen sich Vorstufen des Uriranischen nachweisen. *Anzeiger der Österreichischen Akademie der Wissenschaften, philosophisch-historischen Klasse 120*, 249–55.
- Mayrhofer, M. (1989-2001). *Etymologisches Wörterbuch des Altindoiranischen*. Heidelberg: Carl Winter Universitätsverlag.
- Mayrhofer, M. (1992 [1996]). Archaismen und Neuerungen. Die altpersischen Paralleltexte Dareios Nakš-i Rustam B und Xerxes, Persepolis L. In R. Schmitt (Ed.), *Manfred Mayrhofer: Ausgewählte Kleine Schriften*, pp. 400–9. Wiesbaden: Dr. Ludwig Reichert Verlag.
- McCarus, E. (1958). *A Kurdish Grammar*. New York: American Council of Learned Societies.
- McMahon, A. and R. McMahon (2005). *Language Classification by the Numbers*. Oxford: Oxford University Press.
- Meillet, A. (1922). *Les dialectes indo-européens*. Paris: E. Champion.
- Miller, V. F. (1892). *Materialy dlja izučenija evrejsko-tatskogo jazyka*. St. Petersburg: n.d.
- Monna, M. C. (1978). *The Gathas of Zarathustra: A Reconstruction of the Text*. Amsterdam: Rodopi.
- Morgenstierne, G. (1926). *Report on a linguistic mission to Afghanistan*. Oslo: H. Aschehoug & Co.
- Morgenstierne, G. (1927). *An etymological vocabulary of Pashto*. Oslo: J. Dybwad.
- Morgenstierne, G. (1929). *Parachi and Ormuri*, Volume 1 of *Indo-Iranian Frontier Languages*. Oslo: Instituttet for Sammenlignende Kulturforskning, H. Aschehoug & Co. (W. Nygaard).
- Morgenstierne, G. (1932). Persian etymologies. *Norsk Tidsskrift for Sprogvidenskap 5*, 54–56.
- Morgenstierne, G. (1938). *Iranian Pamir Languages*, Volume 2 of *Indo-Iranian Frontier Languages*. Oslo: H. Aschehoug & Co.
- Morgenstierne, G. (1974). *Etymological Vocabulary of the Shughni Group*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Naṣrī Ašrafī, J. (1381 [2002]). *Farhang-i važiḡān-i Tabarī (A dictionary of Tabari)*. Tehran: “Ṭḡyā’-kitāb”.
- Nakhleh, L., D. Ringe, and T. Warnow (2005). Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language 81*(2), 382–420.

- Nawata, T. (1984). *Mazandarani*, Volume 17 of *Asian and African Grammatical Manual*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Németh, G. (1959). *Eine Wörterliste der Jassen, der ungarländischen Alanen*. Number 4 in *Abhandlungen der deutschen Akademie der Wissenschaften zu Berlin*, Jahrgang 1958. [East] Berlin: Akademie Verlag.
- Nerbonne, J. (2009, January). Data-driven dialectology. *Language and Linguistics Compass* 3(1), 175–198.
- Nerbonne, J., W. Heeringa, E. van den Hout, P. van der Kooij, S. Otten, and W. van de Vis (1996). Phonetic distance between Dutch dialects. In G. Durieux, W. Daelemans, and S. Gillis (Eds.), *CLIN VI, Papers from the Sixth CLIN Meeting*, Antwerp, pp. 185–202. University of Antwerp, Center for Dutch Language and Speech.
- Nichols, J. and T. Warnow (2008). Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2(5), 760–820.
- Olsen, B. (2005). On Iranian dialectal diversity in Armenian. In G. Meiser and O. Hackstein (Eds.), *Sprachkontakt und Sprachwandel: Akten der XI. Fachtagung der Indogermanischen Gesellschaft, Halle an der Saale, 17.–23. September 2000*, Wiesbaden, pp. 473–481. Dr. Ludwig Reichert Verlag.
- Oranskij, I. M. (1963 [1977]). *Les Langues Iraniennees*. Translated by Joyce Blau. Paris: Klincksieck.
- Orel, V. (1995). Slavic *\*ryba* ‘fish’. *Journal of Slavic Linguistics* 3(1), 164–169.
- Parpola, A. (2012). The Dāsas of the *R̥gveda* as Proto-Sakas of the Yaz I-related cultures. With a revised model for the protohistory of Indo-Iranian speakers. In M. E. Huld, K. Jones-Bley, and D. Miller (Eds.), *Archaeology and Language: Indo-European Studies Presented to James P. Mallory*, Number 60 in *Journal of Indo-European Studies Monograph Series*, Washington, D.C., pp. 221–264. Institute for the Study of Man.
- Paul, L. (1998a). The position of Zazaki among West Iranian languages. In N. Sims-Williams (Ed.), *Proceedings of the Third European Conference of Iranian Studies held in Cambridge, 11th to 15th September 1995. Part I: Old and Middle Iranian Studies*, Wiesbaden, pp. 163–177. European Conference of Iranian Studies: Dr. Ludwig Reichert Verlag.
- Paul, L. (1998b). *Zazaki: Grammatik und Versuch einer Dialektologie*, Volume 18 of *Beiträge zur Iranistik*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Paul, L. (2005). The language of the *Šāhnāme* in historical and dialectal perspective. In D. Weber (Ed.), *Languages of Iran: Past and Present: Iranian Studies in Memoriam David Neil MacKenzie*, pp. 141–151. Wiesbaden: Dr. Ludwig Reichert Verlag.



- Paxalina, T. N. (1971). *Sarykol'sko-Russkij Slovar'*. Moscow: Nauka.
- Pelevin, M. (2010). Materials on the Bandari dialect. *Iran and the Caucasus* 14, 57–78.
- Périkhanian, A. (1993). *Materialy k ètimologičeskomu slovarju drevnearmjskogo jazyka*. Yerevan: Izdatel'stvo Nacional'noj Akademij Nauk Respubliki Armenija.
- Pokorny, J. (1959-1969). *Indogermanisches etymologisches Wörterbuch*. Bern: Francke.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rastorgueva, V. S. and D. I. Èdel'man (2000-2003). *Ètimologičeskij slovar' iranskix jazykov*. Moscow: Vostočnaja Literatura.
- Renfrew, C. (2000). At the edge of knowability: Towards a prehistory of languages. *Cambridge Archaeological Journal* 10(1), 7–34.
- Ringe, D., T. Warnow, and A. Taylor (2002). Indo-European and computational cladistics. *Transactions of the Philological Society* 100(1), 59–129.
- Rix, H., M. Kümmel, T. Zehnder, R. Lipp, and B. S. (eds.) (2001). *Lexikon der indogermanischen Verben, die Wurzeln und ihre Primärstammbildungen* (2nd ed.). Wiesbaden: Ludwig Reichert Verlag.
- Ross, M. (1988). *Proto-Oceanic and the Languages of Western Melanesia*. Canberra: Australian National University Press.
- Rossi, A. V. (1998). Ossetic and Baluchi in V.I. Abaev's *Slovar'*. In *Studia Iranica et Alanica: Festschrift for Prof. V.I. Abaev on the Occasion of his 95th birthday*, pp. 373–431. Rome: Istituto italiano per l'Africa e l'Oriente.
- Saitou N. and Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4(4), 406–425.
- Salemann, C. (1901). Mittelpersisch. In W. Geiger and E. Kuhn (Eds.), *Grundriss der iranischen Philologie*, Volume 1, Chapter 3, pp. 249–332. Strassburg: Karl J. Trübner.
- Sapir, E. (1921). *Language: an Introduction to the Study of Speech*. New York: Harcourt, Brace and Company.
- Schapka, U. (1972). *Die persischen Vogelnamen*. Ph. D. thesis, University of Würzburg.
- Schmidt, J. (1872). *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau.
- Schmitt, R. (1989a). Altpersisch. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 56–85. Wiesbaden: Ludwig Reichert Verlag.

- Schmitt, R. (Ed.) (1989b). *Compendium Linguarum Iranicarum*. Wiesbaden: Ludwig Reichert Verlag.
- Schmitt, R. (2000). *Die iranischen Sprachen in Geschichte und Gegenwart*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Schrader, O. (1906-1907). *Sprachvergleichung und Urgeschichte. Linguistisch-historische Beiträge zur Erforschung des indogermanischen Altertums*. Jena: H. Costenoble.
- Schuchardt, H. (1885). *Über die Lautgesetze, gegen die Junggrammatiker*. Berlin: Oppenheim.
- Schwartz, M. (1969). Review of Indo-Scythian Studies, being Khotanese Texts, Vol. VI: Prolexis to the Book of Zambasta by H. W. Bailey. *Journal of the American Oriental Society* 89(2), 444–447.
- Schwartz, M. (1970 [1971]). On the Khwarezmian version of the *muqaddimat al-adab* as edited by Johannes Benzing. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 120, 288–304.
- Schwartz, M. (1974). Irano-Tocharica. In P. Gignoux and A. Tafazzoli (Eds.), *Mémorial Jean de Menasce*, Volume 185 of *Fondation Culturelle Iranienne*, pp. 399–411. Louvain: Imprimerie Orientaliste.
- Schwartz, M. (1982). “Blood” in Sogdian and Old Iranian. In *Monumentum Georg Morgenstierne II*, pp. 189–196. Leiden: Brill.
- Schwartz, M. (2006). On Haoma, and its liturgy in the Gathas. In A. Panaino and A. Piras (Eds.), *Proceedings of the 5th Conference of the Societas Iranologica Europaea*, Volume 1, Milan, pp. 215–224. Mimesis.
- Schwartz, M. (2010). On Rashnu’s scales and the Chinvant’s bridge, with etymological appendices. *Studia Asiatica* 11(1-2), 99–104.
- Schwartz, M. (2012). Loterā’i. *Encyclopædia Iranica* (online version), <http://www.iranicaonline.org/articles/loterai>.
- Séguy, J. (1971). La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35, 335–357.
- Sims-Williams, N. (1989a). Bactrian. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 230–235. Wiesbaden: Ludwig Reichert Verlag.
- Sims-Williams, N. (1989b). East Middle Iranian. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 165–172. Wiesbaden: Ludwig Reichert Verlag.
- Sims-Williams, N. (1989c). Eastern Middle Iranian. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 165–172. Wiesbaden: Dr. Ludwig Reichert Verlag.

- Sims-Williams, N. (1989d). New studies on the verbal system of Old and Middle Iranian. *Bulletin of the School of Oriental and African Studies* 52(2), 255–264.
- Sims-Williams, N. (1989e). Sogdian. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 173–192. Wiesbaden: Ludwig Reichert Verlag.
- Sims-Williams, N. (1996). Eastern Iranian languages. *Encyclopædia Iranica* 7, 649–652.
- Sims-Williams, N. (1998). The Iranian languages. In A. G. Ramat and P. Ramat (Eds.), *The Indo-European languages*, pp. 125–153. London: Taylor & Francis.
- Sims-Williams, N. (2007). *Bactrian documents from Northern Afghanistan, Vol. 2: Letters and Buddhist texts*. London: The Nour Foundation in association with Azimuth Editions.
- Skjærvø, P. O. (1983). *Farnah*:- mot mède en vieux-perse? *Bulletin de la Société de Linguistique de Paris* 78, 241–259.
- Skjærvø, P. O. (1988). Baškardi. *Encyclopædia Iranica* 3(8), 846–850.
- Skjærvø, P. O. (1989). Languages of Southeast Iran. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 363–369. Dr. Ludwig Reichert Verlag.
- Skjærvø, P. O. (2009a). Middle West Iranian. In G. Windfuhr (Ed.), *The Iranian languages*, pp. 196–278. London: Routledge.
- Skjærvø, P. O. (2009b). Old Iranian. In G. Windfuhr (Ed.), *The Iranian languages*, pp. 43–195. London: Routledge.
- Southworth, F. (1964). Family-tree diagrams. *Language* 40(4), 557–565.
- Stanford, J. N. (2012). One size fits all? Dialectometry in a small clan-based indigenous society. *Language Variation and Change* 24, 247–278.
- Steingass, F. J. (1892). *A Comprehensive Persian-English dictionary, including the Arabic words and phrases to be met with in Persian literature*. London: Routledge & K. Paul.
- Sundermann, W. (1989). Parthisch. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 114–137. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Szmrecsanyi, B. (2011). Corpus-based dialectometry: a methodological sketch. *Corpora* 6(1), 45–76.
- Szmrecsanyi, B. (2012). Geography is overrated. In S. Hansen, C. Schwarz, P. Stoeckle, and T. Streck (Eds.), *Dialectological and Folk Dialectological Concepts of Space*, pp. 215–288. Berlin: Walter de Gruyter.
- Tavernier, J. (2007). *Iranica in the Achaemenid period (ca. 550-330 B.C.): lexicon of old Iranian proper names and loanwords*. Leuven: Peeters.

- Tedesco, P. (1921). Dialektologie der westiranischen Turfantexte. *Le Monde Oriental* 15, 184–258.
- Tedesco, P. (1945). Persian *čīz* and Sanskrit *kīm*. *Language* 21(3), 128–141.
- Thomas, B. (1930). *The Kumzari dialect of the Shihuh tribe, Arabia and a vocabulary*. London: The Royal Asiatic society.
- Thordarson, F. (2008 [2009]). *Ossetic Grammatical Studies*. Number 48 in Veröffentlichungen zur Iranistik. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Thurston, W. R. (1994). Renovation and innovation in the languages of North-Western New Britain. In T. Dutton and D. T. Tryon (Eds.), *Language Contact and Change in the Austronesian World*, Number 77 in Trends in Linguistics, pp. 573–610. Berlin: Walter de Gruyter.
- Tischler, J. (1973). *Lexikostatistik und Glottochronologie*, Volume 11 of *Innsbrucker Beiträge zur Sprachwissenschaft*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Toulmin, M. (2009). *From linguistic to sociolinguistic reconstruction: the Kamta historical subgroup of Indo-Aryan*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
- Trask, R. (2000). *The Dictionary of Historical and Comparative Linguistics*. Edinburgh: Edinburgh University Press.
- Tremblay, X. (2005a). Bildeten die iranischen Sprachen ursprünglich eine genetische Familie oder einen Sprachbund innerhalb des indo-iranischen Zweiges? Beiträge zur vergleichenden Grammatik der iranischen Sprachen V. In G. Meiser and O. Hackstein (Eds.), *Sprachkontakt und Sprachwandel: Akten der XI. Fachtagung der Indogermanischen Gesellschaft, Halle an der Saale, 17.–23. September 2000*, Wiesbaden, pp. 673–688. Dr. Ludwig Reichert Verlag.
- Tremblay, X. (2005b). Irano-Tocharica et Tocharo-Iranica. *Bulletin of the School of Oriental and African Studies* 68(3), 421–449.
- Vahman, F. and G. Asatrian (2002). *Notes on the Language and Ethnography of the Zoroastrians of Yazd*, Volume 85 of *Historisk-filosofiske Meddelelser*. Copenhagen: The Royal Danish Academy of Sciences and Letters.
- van Driem, G. (2001). *Languages of the Himalayas: an ethnolinguistic handbook of the greater Himalayan region*. Leiden/Boston: Brill.
- van Etten, J. (2012). R package gdistance: distances and routes on geographic grids.
- Venables, W. and B. Ripley (2002). *Modern Applied Statistics with S* (4 ed.). New York: Springer.

- Vennemann, T. (1974). Sanskrit *ruki* and the concept of a natural class. *Linguistics* 130, 91–97.
- Voskanyan, V. and V. Boyajian-Surenians (2007). Lexical gleanings from Bašākard. *Iran and the Caucasus* 11, 121–125.
- Warnow, T., D. Ringe, and A. Taylor (1995). Reconstructing the evolutionary history of natural languages. IRCS Report 95-16, Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia.
- Watkins, C. (1966). Italo-Celtic revisited. In H. Birnbaum and J. Puhvel (Eds.), *Ancient Indo-European dialects*, pp. 29–50. Berkeley: University of California Press.
- Weiss, M. (2009). *Outline of the Historical and Comparative Grammar of Latin*. Ann Arbor: Beech Stave Press.
- Wendtland, A. (2009). The position of the Pamir languages within East Iranian. *Orientalia Suecana* 58, 172–188.
- Wichmann, S., E. W. Holman, D. Bakker, and C. H. Brown (2010). Evaluating linguistic distance measures. *Physica A* 389, 3632–3639.
- Wieling, M., S. Montemagni, J. Nerbonne, and R. H. Baayen (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and socio-demographic variation using generalized additive mixed modeling. *Language* 90(2), 669–692.
- Wieling, M., J. Nerbonne, and R. H. Baayen (2011). Quantitative social dialectology: explaining linguistic variation geographically and socially. *PLoS ONE* 6(9), e23613.
- Windfuhr, G. (1989). New West Iranian. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 251–262. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Windfuhr, G. (2009). Dialectology and topics. In G. Windfuhr (Ed.), *The Iranian languages*, Chapter 2, pp. 5–42. New York; London: Routledge.
- Yar-shater, E. (1969). *A Grammar of Southern Tati Dialects*. Number 1 in Median Dialect Studies. The Hague: Mouton.
- Zarubin, I. I. (1960). *Šugnanskije teksty i slovar’*. Moscow: Izdatel’stvo Akademij Nauk SSSR.
- Zuraw, K. (2003). Probability in language change. In R. Bod, J. Hay, and S. Jannedy (Eds.), *Probabilistic Linguistics*, pp. 139–176. Cambridge, MA: MIT Press.