Journal of Linguistic Geography (2021), **9**, 106–141
doi:10.1017/jlg.2021.8

CAMBRIDGE UNIVERSITY PRESS

Article

Toward a picture of Chahar Mahal va Bakhtiari Province, Iran, as a linguistic area

Erik Anonby[1,2], Mortaza Taheri-Ardali[3] and Adam Stone[4]

[1]Leiden University Centre for Linguistics (LUCL), Leiden, Netherlands, [2]School of Linguistics and Language Studies (SLaLS) and Department of French, Carleton University, Ottawa, Canada, [3]Department of English and Bakhtiari Studies Research Institute, Shahrekord University, Shahrekord, Iran and [4]School of Linguistics and Language Studies (SLaLS) and Geomatics and Cartographic Research Centre (GCRC), Carleton University, Ottawa, Canada

Abstract

Language documentation has been carried out in Iran since the late 1800s but in a sporadic way, and even now, the scholarly picture of the country's linguistic landscape is fragmentary. The present article responds to this state of affairs in a modest way by working toward a systematic overview of the language situation in one area of the country: Chahar Mahal va Bakhtiari Province of western Iran, where the high Zagros Mountains open onto the Iranian Plateau. In this study, conducted in the context of the *Atlas of the Languages of Iran* (ALI) research programme, we chronicle our research process for this region, beginning with an inventory of languages spoken here—varieties of Bakhtiari, Charmahali, and Turkic—and an overview of their geographical distribution. This initial step enabled us to select 30 varieties from 26 locations across the province for in-depth research, including implementation of the ALI language data questionnaire. Data generated by the study have resulted in two language distribution maps as well as a series of linguistic structure maps. Initial analysis of lexical and phonological data provides insight into defining features of each language as well as structures shared between them as a result of language contact in the region.

Keywords: linguistic geography; language mapping; dialectology; Iran; Chahar Mahal va Bakhtiari Province; Iranic (Iranian) languages; Bakhtiari; Charmahali; Turkic of Iran; Persian

1. Introduction[1]

Chahar Mahal va Bakhtiari Province (hereafter C&B) is nestled in the heights of the Zagros Range in western Iran, with the mountains opening down onto the Iranian Plateau in the north-east. The topography is reflected in the linguistic situation: the Southwestern Iranic language Bakhtiari dominates the mountainous areas that cover most of the province, and two other linguistic groups are intermingled in the lower areas of the north-east: Charmahali, which is also Southwestern Iranic; and Turkic.

C&B is one of Iran's smaller provinces in terms of area as well as population (ISC, 2011/2016) but, as we will show in this paper, it exhibits significant linguistic diversity. However, the character of the province as a linguistic area remains for the most part unstudied. Until 1973, C&B was part of Esfahan Province and—perhaps because of this—the languages of this area were overlooked in the great surveys of the early 20th century (e.g., Mann, 1910; Zhukovsky, 1923; Christensen, 1930, 1935). Even now, Bakhtiari is the only one of the three main varieties that has been documented, and its dialectological characteristics have not been probed in many parts of the province. Existing language maps of the area (TAVO, 1988; Irancarto, 2012; Izady, 2013, among others) have been general and incomplete, and contradict one another.

Author for correspondence: Erik Anonby. Email: e.j.anonby@hum.leidenuniv.nl.

Cite this article: Anonby E, Taheri-Ardali M, and Stone A. (2021) Toward a picture of Chahar Mahal va Bakhtiari Province, Iran, as a linguistic area. *Journal of Linguistic Geography* **9**: 106–141, https://doi.org/10.1017/jlg.2021.8

In this paper, we address this gap in the literature through a first fine-grained and geographically representative study of the language situation in C&B Province. Our research has been conducted as part of the *Atlas of the Languages of Iran* (ALI) research programme (section 2), and it is in this province that our work is most advanced. The paper is divided into two main sections: a detailed description of initial work in bibliographic research and the investigation of language distribution (section 3), both of which were essential in preparing for collection and analysis of linguistic data from 30 language varieties in 26 locations across the province; and exploration of the language situation through analysis of lexicon and phonological correspondences associated with cognate sets (section 4). Preliminary lexicostatistic analyses of the data set the stage for a global understanding of the language situation, and are followed and refined by detailed analysis of individual lexical items.

Our analysis concentrates on the relationship between the two Southwestern Iranic varieties of the province, Bakhtiari and Charmahali, alongside the national language Persian (also Southwestern Iranic). The lexical data show many shared structures among the three Southwestern groups, but several isoglosses distinguish the two regional groups from Persian. Bundling of isoglosses dividing Bakhtiari and Charmahali is even stronger, although several varieties are transitional; and where the larger groups two differ, Charmahali almost always patterns with Persian. In terms of internal linguistic diversity, a putative dichotomy between Rural vs. Urban Charmahali that we observed during the language distribution phase is not borne out by the results of our analysis. In contrast, Bakhtiari shows several clear

dialect areas across the province. In the final section of our analysis, we consider Turkic, concentrating on contact-related patterns of structural similarity between Turkic and the Iranic varieties. While most borrowing is from Iranic (both Persian and the local Southwestern varieties) to Turkic, some structures suggest that Turkic has also made significant contributions to Iranic lexicon and phonology in the region. Results of our study include two language distribution maps and five sample linguistic data maps.

The closing section of this paper re-examines C&B Province as a linguistic area. We bring together salient aspects of the language situation, reflect on limitations in the present study, and identify promising directions for ongoing research.

2. The *Atlas of the Languages of Iran* (ALI) research programme

The current study was conducted in the context of the *Atlas of the Languages of Iran* (ALI) research programme. Initiated in 2009, ALI is now an online, open-access resource (http://iranatlas.net) that is being developed by an international group of institutional partners and scholars.[2] In this section, we provide a summary of the overview written up in Anonby, Taheri-Ardali, and Hayes (2019).

The overall goal of the ALI research programme is to enable work toward a systematic understanding of the language situation in Iran. This initiative, which has the online Atlas at its core, is guided by a set of interrelated themes and questions:

- Linguistic and areal typology: What are important linguistic features of Iran's languages and dialects, and how are they distributed geographically?
- Language distribution: Where are these language varieties spoken, and how does this compare to the distribution of linguistic features?
- Language classification: How do scholars and speakers classify these language varieties, and how can scholarly classifications be improved?
- Language documentation: A record of the linguistic situation in Iran and a repository of linguistic data in the face of declining linguistic diversity with the extension of Standard, Tehran-type varieties of Persian as a mother tongue across the country.

To begin work toward these goals, the Atlas team has reviewed existing efforts to document, classify and map languages of Iran (Taheri-Ardali et al., 2021). An ever-expanding bibliography of linguistic resources[3] is accompanied by a working classification of all language varieties (language families, languages, and dialects).[4] Further, a "multi-dimensional language relation web" has been developed in the Atlas as a way of accounting for competing scholarly classifications and complementary perspectives on language identity, both of which impact the ways in which Atlas users expect language maps to be drawn (Anonby, Hayes & Oikle, 2020; Anonby & Sabethemmatabadi, 2019).

The Atlas is being built using the Nunaliit Atlas Framework (GCRC, 2006–2021), an open-source document-oriented data platform (Hayes et al., 2014; Hayes & Taylor, 2019) that embodies the interactive, multi-modal, and collaborative ethos of the cybercartographic approach to mapping (Taylor, 1997, 2003, 2005; Taylor & Lauriault, 2014; Taylor et al., 2019). Inside a Nunaliit atlas, each piece of data is stored as a document with a flexible set of attributes, and each of these documents can be related to any other document in the atlas. This type of data structure necessitates more initial set-up work in building an atlas, but once an atlas is operational, relations between data are easy to build, navigate, and process. Another key feature of a Nunaliit-designed atlas is its dynamic online platform, which enables direct remote contributions by researchers, and by atlas users generally, from anywhere that has an internet connection, as well as collection and subsequent upload of data from locations without such a connection. To help ensure consistency and reliability, a system for moderation and double-checking of data is an integral part of the data contribution process. In ALI, once data are approved by the editorial team, they are immediately available to Atlas users and are accompanied by clear referencing of the data's source. The Atlas platform therefore serves simultaneously as data repository, collaborative research environment, and publication venue.

Because of the sizable geographic scope of the work, we are proceeding on a province-by-province basis, and further dividing the research into topical areas of activity according to the availability and expertise of Atlas team members. Currently, we have embarked on research for 19 of Iran's 31 provinces, with modest initial results presented and published for six provinces:

- Hormozgan (Mohebbi Bahmani, Rashidi, et al., 2015; Taheri-Ardali, 2017b; Leitner et al., 2021);
- Kordestan (Mohammadirad et al., 2016; Anonby, Mohammadirad & Sheyholislami, 2019);
- Chahar Mahal va Bakhtiari (Taheri-Ardali et al., 2015; Taheri-Ardali, 2017a; Taheri-Ardali & Anonby, 2019);
- Ilam (Gheitasi et al., 2017; Aliakbari et al., 2014; Anonby, Gheitasi & Aliakbari, 2017);
- Bushehr (Nemati, Ghasemi et al., 2017); and
- Kermanshah (Fattahi et al., 2018).

For each province, the Atlas team's first step is assembling a bibliography of documentary studies and scholarly classifications, along with areal overviews and language maps whenever available. This is followed by initial inquiry into language distribution, accompanied by recording of local pronunciations of place names. Results of this preliminary fieldwork are published in the Atlas as province-level language distribution maps, which in turn inform the selection of sites for gathering linguistic data.

Linguistic data is collected by means of a typologically oriented questionnaire designed specifically for the languages of Iran (Anonby, Taheri-Ardali, Haig, et al., 2020; for a detailed description of sources, historical development, content, and justification, see Anonby, Taheri-Ardali & Hayes, 2019:217–20). The ALI questionnaire is divided into four sections: sociolinguistic context, lexicon, morphosyntax, and numbers. A separate section on phonology has now been integrated into the other sections. Instructions for data collection, along with justification for and explanation of the types of linguistic data that the questionnaire aims to gather, are provided as accompanying materials in the *ALI Dataverse* (https://dataverse.scholarsportal.info/dataverse/ali). Published language data are also available there via a permanent link (https://doi.org/10.5683/SP2/FVLDLZ).

The Atlas research process is cyclical, with linguistic data informing earlier findings and refining hypotheses for language classification and language distribution. Of all the provinces of Iran, work on C&B is most advanced, with linguistic data collected from 30 varieties in 26 locations across the province (see 3.1 below). The remainder of this article describes the research conducted there including results generated by preparatory activities (section 3) and an analysis of the lexical questionnaire data, with a

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

focus on the Iranic varieties of the province (section 4). The morphosyntactic questionnaire data as well as data on the province's Turkic varieties are robust and have necessitated separate studies; these are currently being undertaken elsewhere (Anonby, Schreiber & Taheri-Ardali, 2020; Anonby, Taheri-Ardali, Schreiber et al., 2020; Schreiber et al., 2021; Anonby et al., in preparation).

3. Research process for Chahar Mahal va Bakhtiari (C&B) Province

In this section, we describe the process for research we have carried out in C&B Province. We first introduce the research context: the research team and relevant bibliographic materials available to orient the research (3.1). We then provide an overview of the language distribution phase of research (3.2). We present and compare language distribution maps of two types for C&B: an interactive point-based map and a static polygon map. Reflecting on our initial findings, we bring together important research questions related to C&B as a linguistic area (3.3). The activities of language distribution research phase, and in particular the language distribution maps, have facilitated selection of sites for collection of linguistic data using the ALI questionnaire (3.4, 3.5).

3.1 Research context

The research we present here is the result of ongoing work by a large and diverse team. Researchers who contributed to ALI activities for C&B, listed according to their affiliation and specific roles, are as follows:

Mortaza Taheri-Ardali (Shahrekord) *Atlas co-editor, C&B section leader, language distribution, map construction, linguistic data collection and analysis*

Erik Anonby (Carleton/Leiden) *Atlas editor, map construction, linguistic data analysis*

Adam Stone (Carleton) *Map construction, linguistic data analysis*

D.R. Fraser Taylor (Carleton/GCRC) *Project co-investigator*

Amos Hayes (GCRC) *Atlas design, geographic information technology*

J.-P. Fiset (GCRC) *Atlas programming*

Robert Oikle (Carleton/GCRC) *Atlas design, map construction*

Laura Salisbury (Carleton/GCRC) *Map construction*

Mahnaz Talebi-Dastenaee (Alzahra) *Linguistic data collection, map construction*

Peyman Pishyar (Allameh Tabataba'i) *Linguistic data collection*

Maryam Amani-Babadi (Payame Noor) *Bibliography, map construction*

Fatemeh Shahverdi (Tarbiat Modarres) *Linguistic data collection*

Elham Hasanpour (Islamic Azad, Esfahan–Khorasgan) *Linguistic data collection*

Reza Rezvani-Borujeni (Islamic Azad, Khomein) *Linguistic data collection*

In preparation for field research, we first searched out and reviewed existing literature on the languages of the province, including work on their typological features, classification, and geographic distribution. As mentioned in the introduction, C&B Province was passed over in the great linguistic surveys of the early 20th century. To our knowledge, no language maps that focus on the province have been produced prior to the present study. General country-wide language maps such as those of TAVO (Orywal, 1988), Izady (2006–2013), Windfuhr (2009) and Irancarto (Hourcade et al., 2012) indicate three language varieties: Bakhtiari (sometimes subsumed into a larger Lori language grouping), varieties labelled as "Persian," and Turkic. However, the geographic extent shown for each varies greatly. Linguistic work on the province's languages is similarly incomplete: we are not aware of any studies on Persian or Turkic of C&B. Bakhtiari, on the other hand, is more extensively documented. Bibliographies of research on Bakhtiari, covering C&B Province, are found in Anonby & Asadi (2014, 2018) and Anonby & Taheri-Ardali (2018).

Demographic data for C&B have been drawn from the Iranian national census (ISC, 2011, 2016), and settlement-related geographic data are also publicly available on the internet (NCC, 2015; Roostanet, 2016). These data sets provide listings and locations of all populated places, which are needed for language distribution research (3.2) and the subsequent construction of language distribution and linguistic data maps.

3.2 Language distribution: research process

Language distribution research is carried out in ALI for each province by linguists familiar with the language situation. This paper's second author Mortaza Taheri-Ardali organized and led this phase of research for C&B. Taheri-Ardali, born and raised in Ardal in central C&B Province, has worked on Bakhtiari for over a decade and is himself a native speaker of the language. Despite his deep familiarity with the geolinguistic context, the systematic—albeit preliminary—nature of the language distribution research led to a number of new insights, some of which were unexpected. We will discuss these in later sections.

For each of the 840 populated places (cities, towns, and villages) in the census data for C&B, we asked the following basic research questions in relation to language distribution:

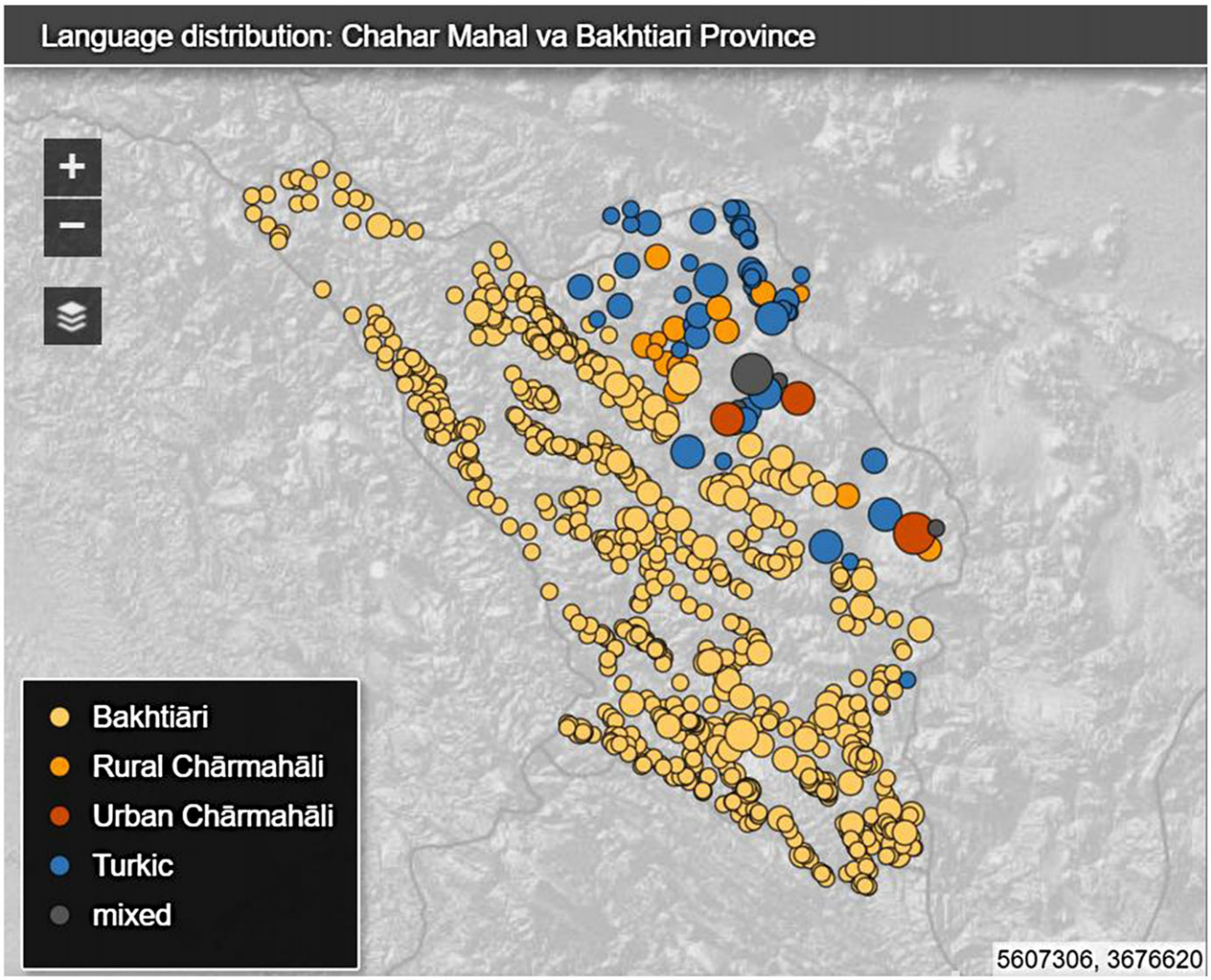
1) What languages, and what subvarieties of these languages, are spoken as a mother tongue in this settlement?
2) In the case that more than one variety is spoken in the settlement, what is the estimated proportion of mother tongue speakers of each variety?

At the same time, on the topic of local place names, we asked: What is/are the local name(s) of this place, as pronounced locally?

Field research on language distribution and local place names was carried out over a 3-month period in early 2015 by Taheri-Ardali, with additional time spent analyzing and verifying the data.[5] The process and results of research are detailed in Taheri-Ardali (2020) and summarized here. Because of logistical difficulties in visiting over 800 settlements, research was undertaken through a network of participants from across the province. The assembled data were based on a convenience sampling of various sources: local knowledge of the field researcher; the field researcher's existing contacts with people from other regions; and additional contacts provided by administrative offices. For advantages and limitations of this sampling approach, as well as its necessity, see the discussion in Anonby, Mohammadirad, and Sheyholislami (2019:16–18, 21–22).

Results from this phase of the research constitute a very general, preliminary approach to the language situation, but they make

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press



Map 1. Interactive map of language distribution in C&B Province. From: http://iranatlas.net/module/language-distribution.chahar_mahal_va_bakhtiari

several key contributions to the overall research process. A first contribution is the place of local knowledge as a starting point complementary to the assessments of experts consulted in the bibliographic research phase. The collection of local pronunciations of place names provides a point of connection for speakers from various regions as potential users of the Atlas. Through transcription and checking of local place names, the field researcher encounters a diversity of language varieties and linguistic structures from across the province. Responses to the language distribution questions allow for a first, fine-grained overview of the language situation in the form of a language distribution map. Here is a screenshot of the resulting interactive point-based map that we have constructed in the Atlas (Map 1):

Due to design-related constraints, only the main mother tongue reported from each community is indicated through colour; and in the case that no single variety is reported as a mother tongue for more than half of the population, the community is simply indicated as "mixed." However, as the following screenshot shows (Map 2), Atlas users can find a full listing of reported varieties by hovering over each point, and more detailed estimates for proportions of each language variety are provided in the side panel.

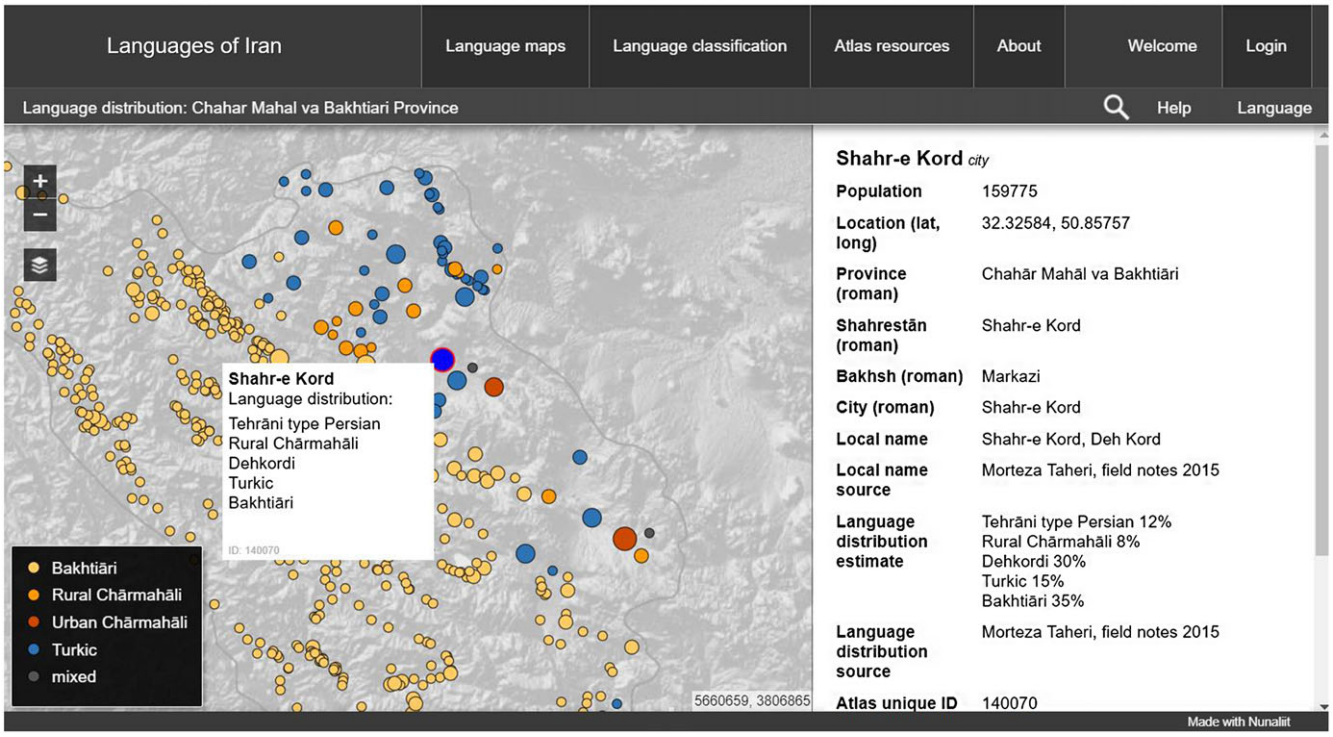
Although the level of detail provided by the map might appear to imply a complete picture of the language situation, we recognize that its assessments are both preliminary and general. This underscores the importance of the Atlas' moderated user contribution feature (section 2 above), where scholars or members of language communities with more detailed or accurate knowledge of the linguistic composition of a given place can confirm or refine the assessment presented in the language distribution map.

An alternative language distribution map, which is static (non-interactive) and uses polygons rather than points to show language distribution, is introduced in 3.4 below as part of the discussion of research site selection.

3.3 Reflecting on language distribution

As mentioned in the research context (3.1) above, existing general language maps of C&B Province show three language varieties: Bakhtiari, Persian, and Turkic. We are not aware of existing published counts or proportions of language communities in the province, but the ethnic composition of the province—often taken as a proxy for linguistic affiliation—is shown in an anonymous infographic in the Persian version of Wikipedia as 56.3%

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press



Map 2. Lists and proportions of language varieties in each place (example). From: http://iranatlas.net/module/language-distribution.chahar_mahal_va_bakhtiari

Bakhtiari, 30.5% Persian, 12.1% Qashqai (a Turkic group), 0.6% "other" and 0.5% unknown (https://fa.wikipedia.org/wiki/استان_چهارمحال_و_بختیاری, accessed 10 July, 2019).

Taheri-Ardali's initial assessment of language varieties in C&B, based on our own language distribution research, presents a significantly different assessment of the language situation. As evident from the map (3.2), we identified four main language groupings in the province: Bakhtiari; Rural Charmahali and Urban Charmahali; and Turkic. Standard-type Persian (defined later in this section) is also spoken and is gaining strength as a mother tongue among all the groups. Formerly, an Armenian language community was also found. Combining estimated language distribution proportions for each settlement with population data from the 2011 census (the latest census data available at the time of initiating research), we calculated the following percentages for number of speakers of each language grouping (Table 1).

Here, we introduce each of the language communities and conclude this section with a set of open questions related to the language situation in the province.

3.3.1 Bakhtiari

Bakhtiari (autoglottonym: *baxtiyāri*), a Southwestern Iranic language with over a million speakers (Anonby & Taheri-Ardali, 2019:445), is the largest and most clearly defined language community in C&B Province. The greater Bakhtiari language area is divided among four provinces of Iran (C&B, Khuzestan, Lorestan, and Esfahan Provinces), but only in C&B does it constitute a linguistic majority. Members of the traditionally nomadic ethnic group who speak this language are found throughout the mountainous western part of the province, and spread down onto the Iranian plateau in the eastern areas. While the lowest areas of the province—in the south-east—are also predominantly Bakhtiari-speaking, the language gives way to Charmahali and Turkic in the north-east corner (see Map 1). There has been a significant migration of Bakhtiari speakers to the capital city of Shahr-e Kord and, with about a third of the population of the city speaking Bakhtiari as a first language, it is now likely the largest mother tongue there.

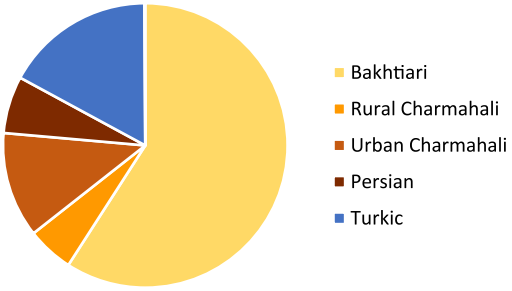
3.3.2 Rural Charmahali and Urban Charmahali

The label "Charmahali" comes from the Persian geographic term *čahār mahāl* (P., lit. 'four regions'), referring to four historical districts in the north-east corner of the province (Lār, Kiār, Mizdej, and Gandomān).[6] Linguistically, this is a heterogenous area, including Turkic-speaking and Bakhtiari-speaking communities as well as other Southwestern Iranic varieties, but the label "Charmahali" can be applied to the Southwestern varieties of C&B Province that are not clearly Bakhtiari or standard-type Persian. Prior to conducting this language distribution research, Taheri-Ardali, who is a native speaker of Bakhtiari, held the (perhaps representative "Bakhtiari") view that Charmahali varieties are essentially a kind of Bakhtiari. On the other hand, speakers of standard-type Persian across Iran, outside of the province, tend to view them as Persian dialects (as noted by Anonby & Sabethemmatabadi, 2019). However, neither perspective is shared by speakers, whose linguistic identity is further subdivided according to a rural vs. urban distinction: Taheri-Ardali notes that in rural areas, speakers refer to their language as Charmahali (autoglottonym: *čārmāhāli*), but in four of the largest cities of the province, speakers refer to their language principally in relation to the name of their city: *dehkordi* (in Shahr-e Kord), *ġafarrokhi* (in Farrokh Shahr), *heyšeguni* (in Hafshejān), and *urǰeni* or *boruǰeni* (in Borujen). Urban speakers feel that their dialects are similar to the urban Southwestern Iranic varieties of Esfahan Province to

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

Table 1. Estimated percentages of mother tongue language speakers in C&B based on ALI language distribution data

Bakhtiari	58.6%
Rural Charmahali	5.4%
Urban Charmahali	12.2%
Turkic	17.3%
Standard-type Persian	6.6%
mixed*	> 0.1%



*The category "mixed" refers to communities or parts of communities that speak various languages and which cannot be further specified without a census of individuals.

the east. While older speakers of Rural and Urban Charmahali varieties see both similarities and differences between their language and Persian, younger speakers—whose language in fact appears to be hybridizing with standard-type Persian—consider themselves speakers of Persian (Taheri-Ardali & Anonby, 2019).

3.3.3 Turkic

The very broad label of "Turkic" is a direct translation of the term *torki*, which local speakers use as the primary point of reference to their own language.[7] But what kind of Turkic? According to Taheri-Ardali (field notes 2015–17), speakers typically respond to this question by referring to the name of their village (for example, "Turkic of Kiān"), and there is little discussion of belonging to any larger dialect grouping within Turkic; for example, they do not view their own variety as a kind of Azerbaijani (*āzeri*), the largest Turkic variety in the country. In the more southerly Turkic-speaking communities of C&B (Sulegān, Boldāji, Naqneh, Juneqān), speakers have an awareness of their historical belonging to the Qashqai tribal confederation of Fars Province, even though there is no longer much contact with this group; in response to the question, "What kind of Turkic?", people from these places answer that they speak Qashqai (autoglottonym: *ġašġāi*) Turkic. However, they do not feel that the Turkic variety they speak is significantly different from varieties spoken in other parts of the province. Further north in C&B, speakers do not identify with Qashqai or with any other subgroup within Turkic, and simply state that they speak Turkic of their own town or village. Some Turkic speakers in various parts of C&B note that the Turkic variety spoken in Ben (one of the linguistic data collection sites; see 3.4 below) is different from the other varieties spoken in the province, but there is no label that unifies or sets apart the "non-Ben" varieties. In the absence of other labels, and pending further analysis of dialect differences among Turkic varieties of the province (Schreiber et al., 2021; Anonby et al., in preparation), we refer to all these varieties as "Turkic of Chahar Mahal va Bakhtiari."

3.3.4 Persian

Standard-type spoken Persian, which brings together elements of *ketābi* (Standard, written) Persian and the Persian dialect of Tehran, and admits varying degrees of substrate influence from regional languages, has entered C&B province as a mother tongue through two channels. First, there is now a sizeable population of immigrants from other provinces of Iran, many of them Persian-speaking, especially in the larger cities of C&B. Even more significantly, Persian is emerging as a mother tongue among existing communities in most areas of the province, as Bakhtiari, Charmahali, and Turkic parents teach Persian to their children as a first language at home. This trend, which began to take shape here about 15 years ago, is most advanced in urban areas. In Taheri-Ardali's (2015) study of multilingualism in the traditionally Bakhtiari city of Ardal, a regional capital of just over 10,000 inhabitants, he estimates that a quarter of the city's population now speaks Persian as a mother tongue. The current generation of children from Bakhtiari families continues to understand the Bakhtiari language and often gains competency in the language through interactions with older relatives and peers outside the home, but Persian is their first language and remains dominant.

3.3.5 Armenian

Until recently, Armenian was spoken in the central-east part of the province, in at least nine towns to the south of Shahr-e Kord: Sirak, Geshniz Jān, Shahrak-e Galugerd, Shalamzār, Qal'eh Mamakā, Boldāji, Ma'mureh, Gandomān, and Vastegān. Most Armenian speakers emigrated to Esfahan during past decades, but they return from time to time to visit the Armenian graveyards in these villages. According to current Charmahali-speaking residents of Sirak, the last Armenian speaker in the village—and possibly in the province—died in 2015 (Mortaza Taheri-Ardali, field notes 2018).

3.3.6 Open questions on the language situation

Several interrelated questions are raised by the assessments that speakers—both inside and outside language communities—as well as "experts" offer in relation to language identification and distribution in C&B, and they need to be addressed before a coherent, stable picture of the language situation can even be put forward. Some of these questions relate more closely to issues of linguistic identity, and others are concerned with genealogical (historical linguistic) relationship or typological similarity.

- What is the relationship of Bakhtiari, Charmahali, and Persian within Southwestern Iranic?
- Should linguists consider Charmahali as a kind of Persian, a kind of Bakhtiari, or even as a distinct language? On what basis?
- In areas where Charmahali and Bakhtiari are spoken alongside each other, is there a clear linguistic distinction between them?
- Are there consistent, defining linguistic differences between Rural Charmahali and Urban Charmahali, or does this putative distinction stem from social perceptions?
- Are there other salient linguistic sub-groupings of Bakhtiari, Charmahali, and Turkic in C&B Province?
- Does the Persian spoken in C&B Province have an areal character, whether as a result of contact with other languages in the province, or as a substrate that shows up among speakers that have shifted from these other languages?

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

Table 2. ALI questionnaire locations for C&B Province

	Language grouping	Location	Site code
1.	Bakhtiari	Sar Āqā Seyyed	SAS
2.	Bakhtiari	Deh Now Soflā	DNS
3.	Bakhtiari	Sepidāneh	SEP
4.	Bakhtiari	Loshtar Gorui	LGI
5.	Bakhtiari	Fārsān	FRS
6.	Bakhtiari	Juneqān	JNB
7.	Bakhtiari	Ardal	ARD
8.	Bakhtiari	Shalamzār	SHL
9.	Bakhtiari	Boldāji	BLB
10.	Bakhtiari	Lordegān	LDG
11.	Bakhtiari	Chilteh Duderā	CHT
12.	Charmahali (Rural)	Cham Chang	CHC
13.	Charmahali (Rural)	Shurāb Kabir	CHK
14.	Charmahali (Rural)	Sheykh Shabān	SHS
15.	Charmahali (Rural)	Arjenak	ARJ
16.	Charmahali (Rural)	Fath Ābād	FTA
17.	Charmahali (Rural)	Hāruni	HAR
18.	Charmahali (Urban)	Shahr-e Kord	SKO
19.	Charmahali (Urban)	Farrokh Shahr	FSH
20.	Charmahali (Urban)	Hafshejān	HAF
21.	Charmahali (Urban)	Borujen	BOR
22.	Charmahali (Rural)	Naqneh	NQC
23.	Turkic	Ben	BEN
24.	Turkic	Margh Malek	MAM
25.	Turkic	Shurāb-e Kabir	SKT
26.	Turkic	(Shahr-e) Kiān	KIN
27.	Turkic	Juneqān	JNT
28.	Turkic	Boldāji	BLT
29.	Turkic	Naqneh	NQT
30.	Turkic	Sulegān	SUL

We will not attempt to answer these questions directly right here, but we will keep them in suspension as we examine linguistic data from across the province, and return to them in the conclusion as a way of informing the initial picture of the language situation that emerges.

3.4 Selection of sites for linguistic data collection

A judicious choice of sites for linguistic data collection using the ALI questionnaire is critical. Each questionnaire interview takes about three hours to carry out, plus travel time; and a single filled-out questionnaire takes several days to transcribe and analyze, along with further time for write-up and construction of linguistic data maps. Although the initial activities leading up to linguistic data collection (3.1-3.3) are also time-intensive, our resulting understanding of the language situation is what facilitates the selection of sites most important to the research goals of the Atlas. Generally, in each *šahrestān* (provincial sub-district) we aim to collect a minimum of one questionnaire in an urban centre, and one in a rural village with limited access to transportation to other areas. This helps provide a representative sample of varieties which are influenced by Standard Persian to different degrees. In districts where our language distribution research (3.2) has identified several language varieties, we plan for collection of the questionnaire in each variety. In some cases, such as for studies of language contact, bilingualism, or generational differences in language structures within a community, we collect more than one questionnaire from the same location. In C&B province, we selected 30 sites (30 varieties in 26 communities), divided among the major language groupings and organized geographically within each grouping in Table 2. Three-letter codes for each site, used in analysis tables later in this article, are also provided.

Subsequent analysis also includes formal Standard Persian (PRS) and Bakhtiari of Masjed Soleymān (MJS) as reference varieties.

As Table 2 shows, 11 questionnaires were collected from Bakhtiari speakers, 11 from Charmahali speakers, including the 4 urban locations where it is spoken (cf. 3.3 above), and 8 from Turkic speakers. In 2 settlements (Juneqān and Boldāji), both Bakhtiari and Turkic questionnaires were collected, and in 2 other settlements (Shurāb Kabir and Naqneh), we gathered Charmahali and Turkic questionnaires.

In addition, for the purposes of inquiry into effects of language contact and bilingualism on linguistic structures where two minority languages co-exist, we collected supplementary questionnaires from second-language speakers of Bakhtiari and Turkic in the town of Juneqān. The first-language data is included in the present study, but we have reserved the second-language data, which is not comparable to the first-language data collected across the province, for description and analysis elsewhere (Anonby, Schreiber & Taheri-Ardali, 2020).

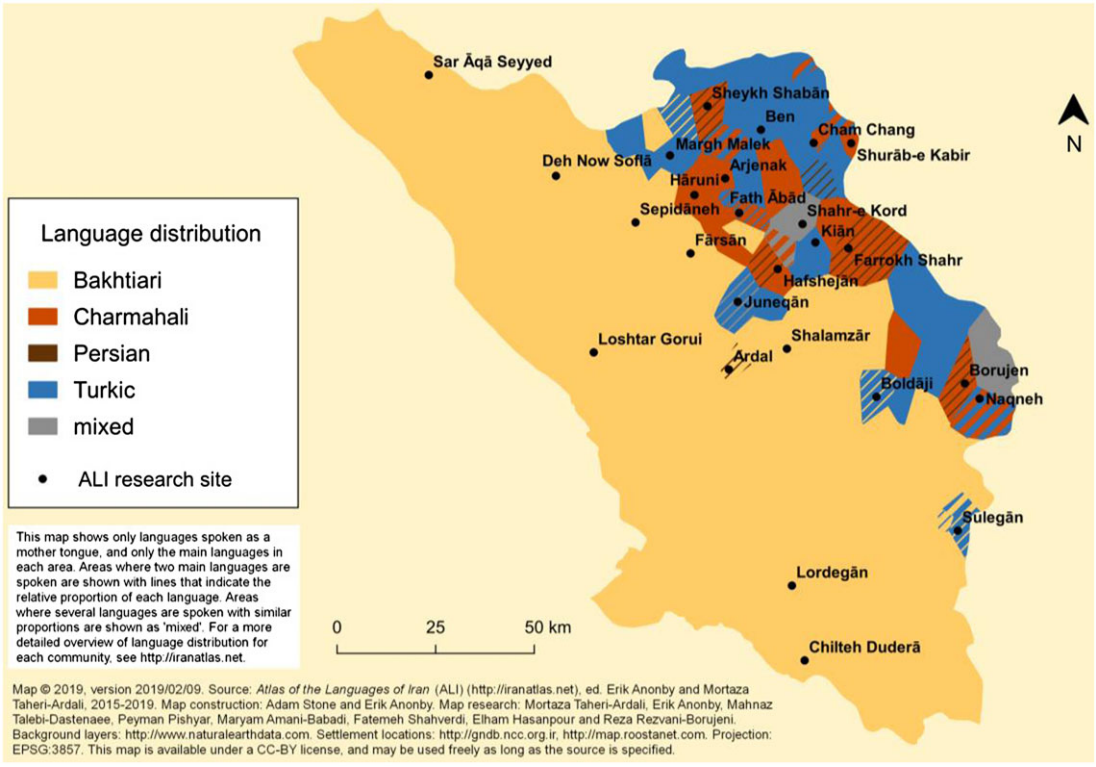
ALI research sites for C&B province are summarized on the following map (Map 3), with a simplified static polygon representation of language distribution as a backdrop. This map is based on the same language distribution data as the interactive point-based map above (3.2), but spaces between settlements have been filled in using contiguous "Voronoi" polygons (see Burrough, McDonnell & Lloyd, 2015:160–63). In addition, pending a more definitive analysis of the relationship between Rural and Urban Charmahali (see section 5), we have combined them and represented them with a single colour for the purposes of this map. Although the representation here is not interactive, it has the advantage of being able to show more than one language in a single settlement, and in relative proportions. This static polygon map further provides a simpler backdrop ideal for presenting the ALI research sites for C&B.

3.5 Collecting linguistic data with the ALI questionnaire

Over a two-year period from 2015 to 2017, Taheri-Ardali led work in collecting linguistic data by carrying out the ALI questionnaire, sometimes in conjunction with other ALI team members, in each of these 30 sites. All of the questionnaire sessions were recorded, and the sound files are being prepared for upload to interactive linguistic data maps in the Atlas. Full research metadata and linguistic data are archived and available from a permanent link (https://doi.org/10.5683/SP2/FVLDLZ).

Introduced in section 2 above, the questionnaire carried out for this study contained five sections: sociolinguistic context, lexicon, phonology, morphosyntax, and numbers. The sociolinguistic

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press



Map 3. ALI research site selection for C&B. From: http://iranatlas.net/module/language-distribution.chahar_mahal_va_bakhtiari_static

context section of the questionnaire provides a much fuller picture of the linguistic situation at the research location than the very cursory information collected in the earlier language distribution research phase (3.2 above); here, we also have the opportunity to ask questions about ethnicity vs. language identity, second/additional languages, fluency in Persian across the community, and language shift and endangerment.

The lexicon section of the questionnaire, which will be the focus of the analysis in the remainder of the present study, contains 80 words. These words have been selected based on their inclusion in other important wordlists and studies (Swadesh, 1971; Leipzig-Jakarta [Haspelmath & Tadmor, 2009]; Persian Academy, 2009) as well as "classical" Iranic isoglosses (e.g., Oranskij, 1979; Schmitt, 1989) and other areal patterns that scholars have identified while working on languages of Iran (e.g., Stilo, 2016). As we will show, analysis of the geographic distribution of lexical items is important for understanding the language situation (3.2, 3.3), but analysis of the distribution of shared phonological forms among cognates is also valuable for assessing patterns of change and diffusion (section 4). The 2400 lexical data items we collected and analyzed—the 80 questionnaire items from each of the 30 research locations—are inventoried in Appendix 1, along with lists of 80 items for two reference varieties: formal Standard Persian of Tehran, and Bakhtiari of Masjed Soleymān in Khuzestan Province. While this article provides general discussion on the Turkic lexicon, we focus on the Iranic varieties of the province: Bakhtiari, and Rural and Urban Charmahali.

Because of the richness of the patterns we have observed in the lexical data in their own right, the phonology, morphosyntax and numbers sections of the questionnaire are beyond the scope of this study. The morphosyntactic data are certainly complementary and of equal importance in understanding C&B as a linguistic area, but they are even richer and more complex than the lexical data, and they need to be treated in one, or perhaps several, additional dedicated articles. Topics of special typological interest that we have observed in the morphosyntactic data are plural marking, definiteness, differential object marking, person marking, and noun phrase structure.

As mentioned just above, the lexical data are relevant for an understanding of some aspects of phonology, particularly in relation to its historical and geographic patterning. In contrast, because of the system- and discourse-dependent nature of phonology, responses to the phonology section of the questionnaire did not enable us to establish a satisfactorily coherent or comprehensive picture of the topics we investigated: consonant, vowel, and diphthong inventories, and dominant stress and intonation patterns, among others.

The data we collected on numbers, based on Chan's (2008–2019) questionnaire, are much clearer. As a homogeneous set within the lexicon and susceptibility to borrowing, they also constitute a cohesive topic on their own; we will therefore also treat them in a separate study.

Subsequent to the field research carried out for the present study as well as field research at 20 other locations across Iran, a group of scholars met in Bamberg, Germany in July 2017 to test and revise the questionnaire.[8] Going forward, the improvements made possible through this process will further ground the results of our research.

4. Analysis of linguistic data

In this section, which forms the second major thrust of the paper, we analyze and compare lexical data from Bakhtiari (henceforth B.) and Charmahali (Ch.) as well as Turkic (T.) varieties across the province. The archived data set (Taheri-Ardali, Anonby et al., 2020) comprises full research metadata plus 80 lexical items from the 30 research locations introduced above (3.5). These 2400 lexical data items are

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

catalogued in Appendix 1, along with lists of 80 items from Standard Persian as well as Bakhtiari of Masjed Soleymān as reference varieties. We also refer to colloquial Persian in the analysis when relevant. Each of the sound correspondences discussed in this section is catalogued in Appendix 2.

In order to provide a general impression of relative structural similarity between and within these main linguistic groupings, our analysis begins with lexicostatistic analysis using calculations of cognate percentages between pairs of lexical data sets (4.1). As we will point out, these measures of similarity are in many ways limited. The larger part of our analysis therefore concentrates on identifying recurrent distributional patterning of lexical and phonological structures, and seeks to account for them as part of a meaningful areal picture of linguistic contact, change, and diffusion. Analysis is concentrated on the Iranic varieties B. and Ch., which cover most of the province, and attempts to address questions of their debated relation to each other as well as the internal variability of each. We first affirm the place of B. and Ch. alongside Persian within the Southwestern group of the Iranic family (4.2). We then look at linguistic structures common to the Iranic languages of C&B but not found in Standard Persian (4.3). Isoglosses distinguishing B. and Ch. are identified (4.4) and compared with the local perceptions of language identification introduced above (3.3), and tendencies in the geographic distribution of linguistic structures across the B. language area in particular are delineated (4.6). Tabulation and discussion of shared sound correspondences provides a finer-grained, summative overview of patterns of similarity among Iranic varieties (4.7). Finally, we consider data from Turkic varieties alongside the Iranic varieties (4.8). This final component, albeit cursory, provides insight into the regional patterning of language contact and helps to complete the picture of C&B as a linguistic area.

4.1 Lexicostatistic analysis

Lexicostatistic analysis measures structural similarity between lexical items with equivalent meanings in different language varieties. The most common metric for calculating lexical similarity is by counting the proportion of cognates—that is, sets of related words—for pairs of language varieties (early examples include Swadesh, 1950 and Dyen, 1962). Cognates can be established through prior historical-comparative analysis or, in cases where such analysis has not been carried out or is not intended, through the simpler but more subjective measure of apparent similarity. Although a high number of cognates is often taken as an indicator of close relation between languages, there are other reasons that languages may share cognates, such as borrowing between languages and universal tendencies for some lexical domains. Further, lexicon is only one component of language, and each particular list of lexical meanings will generate a different proportion of similar items. While the lexical items gathered in this study are based in part on well-known tools such as the 100-item Swadesh (1971) and Leipzig-Jakarta (Haspelmath & Tadmor, 2009) lists, other words have been included in light of their importance for the Iranian linguistic context (3.5). For these reasons, the cognate percentages generated here should be taken only as a relative indicator of similarity, and relevant for this particular set of words, as a means of highlighting general tendencies that will be refined through detailed analysis in the following sections of this study.

Using the program Wordsurv (http://lingtransoft.info/apps/wordsurv), we grouped words into cognate sets and tabulated shared cognate percentages for the 30 language varieties treated in this article, along with Persian and Bakhtiari of Masjed Soleymān as reference varieties. In the following table (Table 3), B. of Masjed Soleymān (MJS, top row) is followed by 11 other B. varieties from Sar Āqā Seyyed (SAS) in the north-west to Chilteh Duderā (CHT) in the south. Ch. varieties follow a similar progression across the north-east part of the province from Cham Chang (CHC) in the north to Naqneh (NQC) to the south-east. Persian (PRS) is placed after the Ch. varieties, with which it shares many structural similarities. T. varieties run parallel to Ch. varieties in the north-east corner, running from Ben (BEN) in the north to Sulegan (SUL) in the south-east.

Percentage of shared cognates between each of the 32 language varieties are as follows (Table 3).[9]

Several general lexicostatistic patterns are evident from Table 3.

1. Most obviously, Iranic and Turkic varieties are clearly differentiated by their vocabulary. The highest percentage of cognates shared by the two families, almost all of which are attributable to borrowing of Iranic words into Turkic (4.8), is found between T. of Shahr-e Kiān (SHK) and two Ch. sites (22%); there are six other T.–Ch. wordlist pairs that have a similar level of 21% lexical similarity. At the other end, the lowest proportion of shared cognates between the two families, between T. of Ben (BEN) and B. of Fārsān (FRS), is just 5%. T. of Sulegān (SUL) and Naqneh (NQT) also show low levels of similarity with Iranic vocabulary. There is a general, but modest, trend of higher levels of similarity between T. and Ch. (which are more often geographically proximate) than between T. and B. varieties.

2. Ch. varieties show a high level of lexical similarity to Persian (with one exception, above 90%), but B. varieties show varied levels ranging from 65% to 86%. The B. locations with the lowest levels of similarity to Persian are found at the north-west and south ends of the province (north-west: SAS, DNS; south: LDG, CHT).

3. Ch. varieties show a high level of lexical homogeneity. The words collected from Shurāb Kabir Ch. (SKC) are cognate with 100% of wordlist items from three other Ch. sites.[10] The lowest level of pairwise lexical similarity between two Ch. varieties, Sheykh Shabān (SHS) and Naqneh Ch. (NQC), is a still relatively high value of 86%.

4. B. varieties show varying levels of lexical similarity with one another. As for a few pairs of Ch. varieties, where all wordlist items exhibit cognates, 100% of items from the B. communities of Ardal (ARD) and Sepidāneh (SEP) share cognates. In contrast, B. of Chilteh Duderā (CHT) in the far south and Boldāji (BLB) in the west share cognates for only 76% of items.

5. The patterning of lexical similarity between Ch. and B. shows an interesting areal trend. B. varieties at the north-west and south ends of the province, which show the lowest levels of similarity to Persian (point 2 above), also show relatively low levels of similarity to Ch. varieties. However, B. varieties in the centre of the province, generally in close proximity to Ch. and T. varieties, are in general quite similar to Ch. In fact, B. of Shalamzār (SHL) and Boldāji (BLB) both have minimally 85% lexical similarity with *all* Ch. varieties in the sample. This pattern will be explored further at various points in the paper (see especially 4.6 4.7).

6. The T. varieties in the data show a fairly high level of lexical similarity, ranging from 78% shared cognates between T. of Juneqān (JNT) and two other locations, to 95% between Ben (BEN) and Margh Malek (MAM). The modest sample of 8 T. varieties does not show any clear geographic pattern of areal variation in the lexical similarity counts.

As underscored above, these results of lexicostatistic analysis are limited to helping sketch out a preliminary initial picture of the language situation based on relative similarities between varieties in a single structural domain (lexicon). However, they have enabled the

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

Table 3. Percentages of shared cognates in the 80-item wordlist

	MJS	SAS	DNS	SEP	LGI	ARD	FRS	JNB	SHL	BLB	LDG	CHT	CHC	SKC	SHS	ARJ	FTA	HAR	SKO	FSH	HAF	BOR	NQC	PRS	BEN	MAM	SKT	KIN	JNT	BLT	NQT	SUL
MJS	100	91	94	92	91	92	91	85	88	84	89	85	74	79	81	80	76	79	78	76	78	80	74	70	6	12	14	16	16	14	5	9
SAS	91	100	96	92	92	94	90	88	88	82	92	84	74	78	82	79	76	80	78	76	75	80	70	74	6	15	14	16	19	15	8	10
DNS	94	96	100	92	94	92	90	89	88	82	92	86	75	80	81	81	79	82	78	78	78	79	75	74	6	15	14	18	19	16	8	10
SEP	92	92	92	100	96	100	99	92	96	90	94	86	80	85	89	86	82	88	84	84	82	88	78	79	6	14	15	19	18	15	6	10
LGI	91	92	94	96	100	96	96	91	92	89	95	88	79	82	88	82	81	84	81	81	80	84	75	78	6	15	15	18	19	15	6	9
ARD	92	94	92	100	96	100	99	96	96	92	94	86	84	88	91	88	86	88	86	86	85	89	81	81	8	16	16	18	20	16	9	11
FRS	91	90	90	99	96	99	100	94	95	92	92	85	81	85	89	85	84	85	84	84	82	86	78	79	5	14	14	18	18	14	8	9
JNB	85	88	89	92	91	96	94	100	96	94	88	82	85	90	91	88	91	91	88	88	86	88	81	84	10	18	19	20	21	20	11	14
SHL	88	88	88	96	92	96	95	96	100	98	90	81	89	94	94	92	94	94	91	92	88	94	86	85	8	16	16	19	20	18	9	11
BLB	84	82	82	90	89	92	92	94	98	100	85	76	90	95	92	92	94	95	92	94	88	94	88	86	8	15	16	19	18	18	10	11
LDG	89	92	92	94	95	94	92	88	90	85	100	98	75	80	85	80	78	81	78	79	78	81	76	74	9	16	18	19	19	18	8	11
CHT	85	84	86	86	88	86	85	82	81	76	98	100	69	72	76	74	71	74	71	70	72	72	69	65	10	18	16	19	20	19	9	11
CHC	74	74	75	80	79	84	81	85	89	90	75	69	100	99	89	96	95	95	98	95	96	98	91	95	8	16	19	20	20	18	11	14
SKC	79	78	80	85	82	88	85	90	94	95	80	72	99	100	92	99	99	99	100	100	98	100	95	95	9	16	19	20	20	19	11	14
SHS	81	82	81	89	88	91	89	91	94	92	85	76	89	92	100	90	91	90	91	94	86	92	86	84	10	18	21	22	21	21	12	16
ARJ	80	79	81	86	82	88	85	88	92	92	80	74	96	99	90	100	96	96	96	96	94	98	91	90	9	16	18	20	20	19	10	12
FTA	76	76	79	82	81	86	84	91	94	94	78	71	95	99	91	96	100	98	98	96	92	95	91	90	10	18	19	21	20	20	11	14
HAR	79	80	82	88	84	88	85	91	94	95	81	74	95	99	90	96	98	100	98	96	95	96	92	95	10	18	19	22	20	20	11	14
SKO	78	78	78	84	81	86	84	88	91	92	78	71	98	100	91	96	98	98	100	96	98	98	94	95	9	18	21	21	20	20	11	14
FSH	76	76	78	84	81	86	84	88	92	94	79	70	95	100	94	96	96	96	96	100	91	99	92	90	8	15	20	20	19	20	10	15
HAF	78	75	78	82	80	85	82	86	88	88	78	72	96	98	86	94	92	95	98	91	100	94	91	91	8	16	19	19	20	18	11	14
BOR	80	80	79	88	84	89	86	88	94	94	81	72	98	100	92	98	95	96	98	99	94	100	94	92	8	15	19	20	20	19	10	15
NQC	74	70	75	78	75	81	78	81	86	88	76	69	91	95	86	91	91	92	94	92	91	94	100	90	6	12	16	18	16	16	9	14
PRS	70	74	74	79	78	81	79	84	85	86	74	65	95	95	84	90	90	95	95	90	91	92	90	100	9	18	19	18	19	16	12	14
BEN	6	6	6	6	6	8	5	10	8	8	9	10	8	9	10	9	10	10	9	8	8	8	6	9	100	95	81	85	78	89	90	90
MAM	12	15	15	14	15	16	14	18	16	15	16	18	16	16	18	16	18	18	18	15	16	15	12	18	95	100	86	88	85	95	94	91
SKT	14	14	14	15	15	16	14	19	16	16	18	16	19	19	21	18	19	19	21	20	19	19	16	19	81	86	100	81	78	85	81	81
KIN	16	16	18	19	18	18	18	20	19	19	19	19	20	20	22	20	21	22	21	20	19	20	18	18	85	88	81	100	88	90	80	82
JNT	16	19	19	18	19	20	18	21	20	18	19	20	20	20	21	20	20	20	20	19	20	20	16	19	78	85	78	88	100	86	79	81
BLT	14	15	16	15	15	16	14	20	18	18	18	19	18	19	21	19	20	20	20	20	18	19	16	16	89	95	85	90	86	100	85	88
NQT	5	8	8	6	6	9	8	11	9	10	8	9	11	11	12	10	11	11	11	10	11	10	9	12	90	94	81	80	79	85	100	94
SUL	9	10	10	10	9	11	9	14	11	11	11	11	14	14	16	12	14	14	14	15	14	15	14	14	90	91	81	82	81	88	94	100

(place names represented here by 3-letter research location codes are defined in Table 2 above)

identification of several patterns of varying clarity and significance. Subsequent analysis (4.2–4.8), which focuses on individual data items, and moves from lexical similarity to analysis of sound correspondences, allows for the evaluation and refinement of these patterns.

4.2 Structures common to Bakhtiari, Charmahali, and Persian

Because of common Southwestern Iranic ancestry, B. and Ch. share many structures with each other and with Persian. In fact, of the 80 lexical items in the ALI questionnaire, 5 are identical across the 11 B. locations, 11 Ch. locations, and Persian: 4. *guš* 'ear,' 15. *pā* 'foot,' 34. *gorg* 'wolf,' 44. *ruz* 'day,' and 52. *gerdu* 'walnut.' This pattern is shown in Map 4 for *gorg* 'wolf,' where the Iranic items are uniform but the Turkic equivalents are everywhere *ġurt*. Not only are these lexical items cognate across the Iranic varieties examined; they are also phonologically uniform as a result of shared sound changes. Together, this points to a shared historical origin and path, and underlines the relative proximity of the genealogical relationship between B., Ch., and Persian.

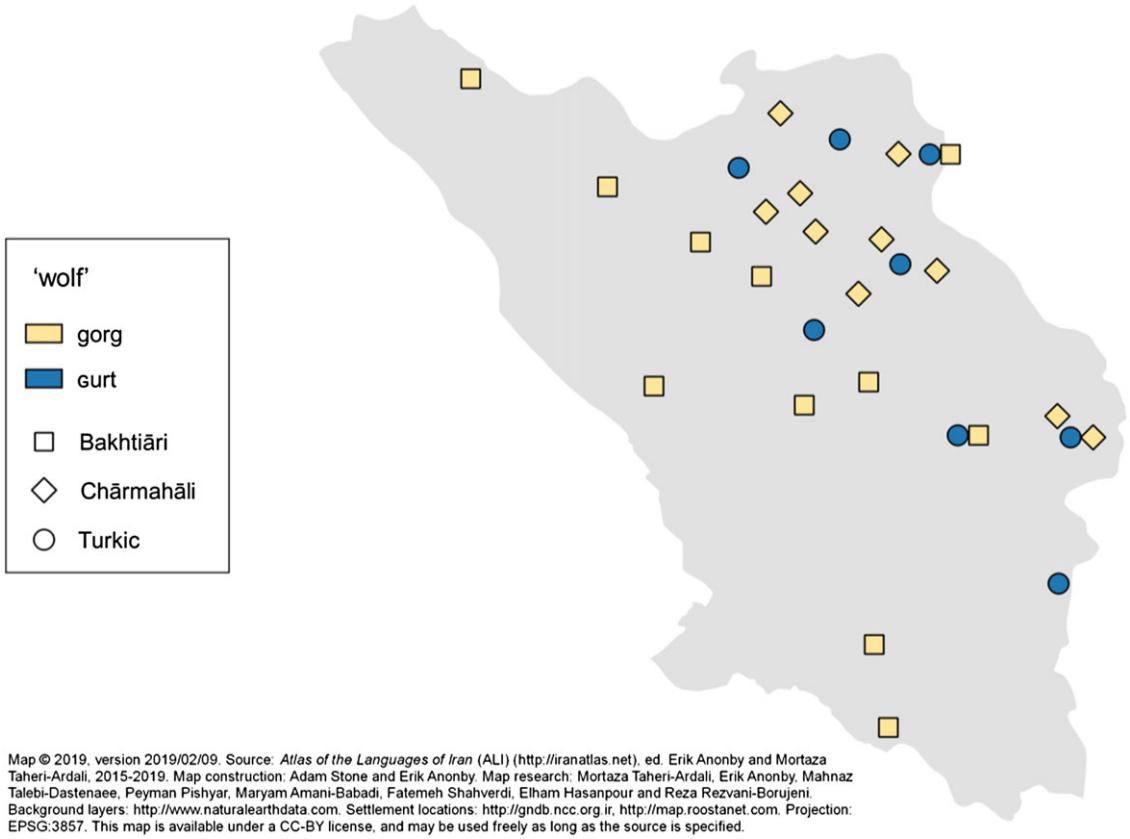
For a further 5 items, the words are generally uniform in all questionnaire data sets and in Persian:

- 8. 'arm' (P. *dast*) and 9. 'hand' (P. *dast*): these items, for which the semantic delimitation and distinction is challenging in the Iranian context, are served by a single word in most given locations, but the variants *dast* and *das* appear with equivalent frequency across both B. and Ch. language areas (exceptions: *dahs* [da:s][11] in the most southerly research location of Chilteh Duderā (B.), as well as distinctive lexical items *čel* and *bāyi* 'arm' and *panga* 'hand' elicited in the B. villages of the far north-west);
- 13. 'leg' (P. *pā*), which is *pā* in most varieties (parallel to 15. *pā* 'foot,' mentioned in this section above), but given as *leng* in a handful of both B. and Ch. locations, and *tekerav*/*tekerun* in peripheral B. sites (regarding a B. "periphery," see 4.6 below);
- 48. 'winter' (P. *zemestān*), where B. and Ch. are always cognate with the Persian equivalent, but with considerable variation in vowels (first syllable *ze* ~ *za*; second syllable *me* ~ *ma* ~ *meh* [mɛ:]), word-internal *st* ~ *ss* ~ *s*, and word-final codas (usually *un* [ũ:]/[ũn], but also *on* [õ:]/[õn], *ovn* [õw̃], *avn* [ɔ̃w̃], *av* [ɔw] and *u* [u:]); and
- 54. 'thirsty' (P. *tešne*), given everywhere as *tešne*, except for the most southerly location of Chilteh Duderā (B.), where the word-final support vowel is different (*tešna*).

Further cases of cognacy, but for which sound changes and correspondences pattern significantly between varieties or over geographic areas, are treated in the relevant sections below (4.3–4.8).

One typically Southwestern sound change common to B., Ch., and most spoken varieties of Persian (but not found in formal registers of Standard Persian), is historical raising of *ā* (usually to *u*) before a nasal *n* or *m*. Although the exact shape of the affected

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press



Map 4. Lexical variation in C&B Province: 'wolf.' From: http://iranatlas.net/module/linguistic-data.cb-lexicon-wolf

lexical items is highly variable, the sound change applies consistently. Example items are found in Table 4.

For the word 14. 'knee,' (P. *zānu*) which (for whatever reason) has not undergone historical raising in colloquial Persian (coll. P. also *zānu*), the vowel is raised in 17 of the 22 Iranic varieties in the sample: *zuni* is the most common form of the word there.

For 5. 'mouth,' there are two variants in Persian: *dahan*, and *dahān* (the latter of which is more formal). Interestingly, Ch. (*dahan*, *dahn*, *dan*) patterns with the first variant in 10 of 11 locations, and B. (*dohun*, *dun*, etc.) exhibits a form closer to the second variant in all of the 10 locations where a cognate is found.

4.3 Bakhtiari and Charmahali together against Persian

Structural similarities found between Bakhtiari and Charmahali, but not shared by Persian, suggest close relation between the region's two Iranic groups, a pattern of areal borrowing and convergence, or both.

Regionally distinctive vocabulary. A couple of regionally distinctive words in the lexical data are shared by B. and Ch., but different from Persian:

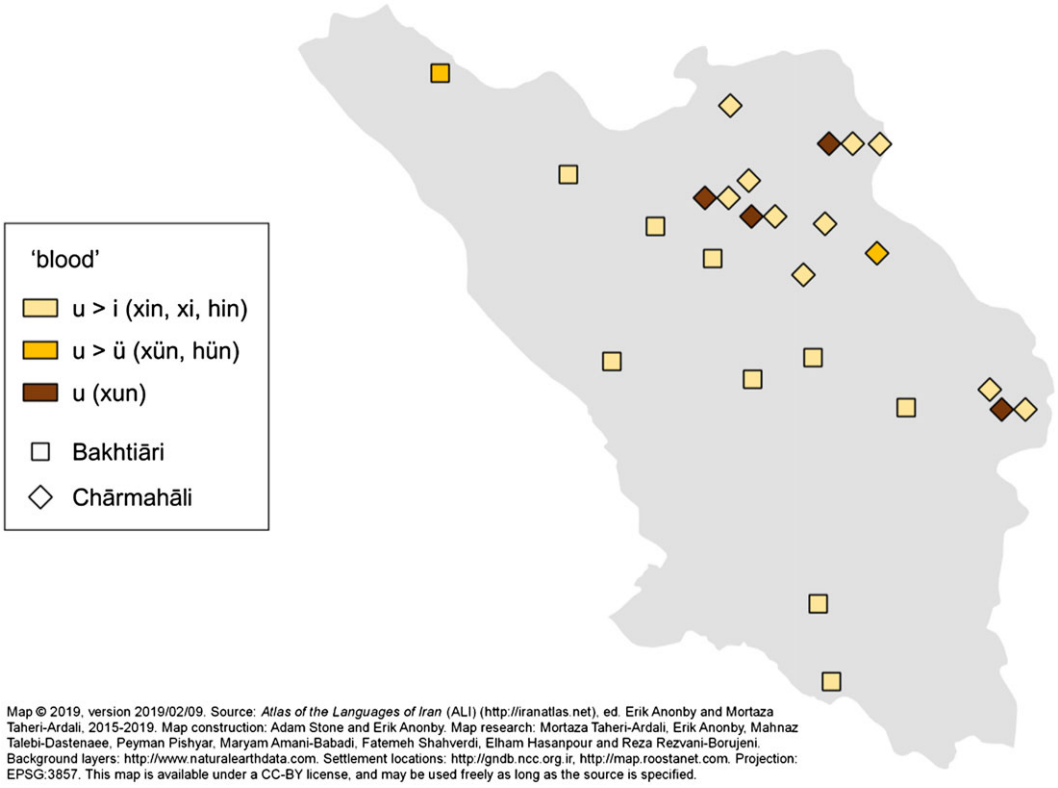
- 72. 's/he swept' (P. *ǰāru kard*)[13] appears as *roft* (especially B.) or *ruft* (especially Ch.) alongside or in place of *ǰāru kerd* (or similar) in all but 5 of the 22 Iranic locations.
- 79. 'tomorrow')P. *fardā*) is found as variants of the word *sobā*/*sovā*/*sovah* (cf. Ar. *sabh* 'morning') and its cognate *sob*/*sov*/*sohv* (cf. Ar. via P. *sobh* 'morning') in all B. and Ch. locations, although *fardā* is found as a doublet in three Ch. locations. (For geographic distribution of *b* vs *v* in this item, which is significant, see 4.4.) A similar areal pattern exists for the lexical item 80. 'day after tomorrow' (P. *pasfardā*), for which variants of *passobā*, *passova*, *pahsovah*, etc., are used in all Iranic varieties of the region.

Table 4. Words exhibiting historical *ā* > *u* before nasals in colloquial Persian, Charmahali and Bakhtiari

	Standard Persian	Colloquial Persian	Charmahali	Bakhtiari
6. 'tongue'	*zabān*	*zabun*	*zebun, zobun, ...*[12]	*zevun, zun, ...*
16. 'bone'	*ostoxān*	*ostoxun*	*ostoxun, estexun, ...*	*ostoxun, ...*
30. 'groom'	*dāmād*	*dumād*	*dumād, ...*	*dumā, ...*
48. 'winter'	*zemestān*	*zemestun*	*zemestun, zemastun, ...*	*zemestun, ...*
49. 'house'	*xāne*	*xune*	*xune, ...*	*hune, ...*
62. 's/he came'	*āmad*	*umad*	*umad, ...*	*uma, omey, ...*
66. 's/he knew'	*midānest*	*midunest*	*idunest, midunest, ...*	*idunest, ...*
74. 'there'	*ānǰā*	*unǰā*	*unǰā, ...*	*unǰā, učo, očo, ...*

In addition, three recurrent phonological isoglosses (i.e., isophones) distinguish all of the region's Iranic varieties from

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press



Map 5. Phonological variation in Iranic of C&B Province: Historical fronting of *u* in *xun* 'blood.' From: http://iranatlas.net/module/linguistic-data.cb-historical-phonology

Standard Persian: historical fronting of *u*, retention of *i* in open word-initial syllables of selected words, and historical softening of *b* in codas.

Historical fronting of u. The first of these sound changes is the historical fronting of *u* to *i* (or sometimes only to the intermediate form *ü* [y]), generally in the environment of coronals. This fronting has taken place in 1. 'hair' (P. *mu*), which is found as *mi* or *mü* in 17 of the 19 Iranic locations where a cognate item is used; 7. 'throat' (P. *galu, gelu*), found with a fronted vowel (*gili*, *gülü*, *gelü*) in 19 of the 22 Iranic locations; 14. 'knee' (P. *zānu*), with a final fronted vowel (*zuni*, *zāni*, *zavu*, *zuvi*) in 20 of 22 sites; and 17. 'blood' (P. *xun*), found with a fronted vowel (*xin*, *hin*, *xün*, *hün*) in all 22 locations, but sometimes alongside the non-fronted form characteristic of Persian (see Map 5).

For all four of these words, doublets identical to Persian are also in use in a handful of locations; these locations are not consistent, but parallel or exclusive use of non-fronted (=Persian) forms is more common in Ch., and in the Ch. variety of the city of Shahr-e Kord in particular.

The historical fronting process appears to have been active into post-*hamleh* history (that is, subsequent to the Arab conquest of Persia in the 7th century), applying to the Arabic borrowing 31. 'bride' (Ar. *ʕarūs* > P. *arus*) in a few Ch. and B. locations, where the form *āris* is used.

Even though 44. 'day' (P. *ruz*) exhibits a phonological environment where *u* fronting would be expected to take place, it is uniformly *ruz* in the Iranic varieties in the sample. This could be due to the presence of a different historical vowel *ō* (cf. Middle Persian [=MP] *rōz*) when the sound change was taking place; a result of avoiding confusion with the existing Southwestern Iranic word *riz* 'fine, tiny'; or the prominence of the Persian word for 'day' in shared cultural contexts. Interestingly, 77. 'yesterday' (P. *diruz*) undergoes fronting in 11 of the 13 locations where a (partial) cognate is used (*diriz*, *dürüz*, *periz*[14]), as does 78. 'day before yesterday' (P. *pariruz*), which shows a fronted vowel in 6 of the 18 cognate forms (*paririz*, *peririz*). This inconsistent historical application of the sound change suggests that phonological conditioning is only partially determinative, and that additional factors such as lexical semantics or contact-related influence from Persian are also significant for sound changes.

Retention of i in word-initial open syllables. Another isogloss distinguishing Iranic varieties of C&B from (Standard or Tehrani-type) Persian is the presence of *i* in word-initial open syllables in four words in the C&B data where Persian has *e*: 7. 'throat' (P. *galu*, *gelu*) is *gili* in 10 of 11 B. varieties and exhibits varied forms with an initial high vowel (*gili*, *gilu*, *gülü*) in all Ch. locations; for 11. 'stomach' (P. *šekam*), B. varieties do not contain *i*, ostensibly due to the different shape of the cognate (*ešgam*, *eškam*, *kom*) (see 4.4), but all 9 of the 11 Ch. varieties with a cognate item exhibit the form *šikam*; in 20. 'liver' (Standard P. *ǰegar*, coll. P. *ǰigar*), a high vowel is used in 9 of 11 B. locations (*ǰiyar, ǰigar*), and all 11 Ch. locations (*ǰigar*); and in 61. 'white' (P. *sefid*), the first vowel is *i* in all Ch. locations (again, most B. varieties have a different shape here—*esbid*, etc.—and the first vowel is non-high). As Agnes Korn (pers. comm. 2020) has pointed out, it is in fact Standard/Tehrani-type Persian which has undergone a broader, unconditioned change in vowel quality (*i* > *e*) here, whereas Iranic varieties of C&B retain a non-lowered historical *i* in these words. The vowel *e* is, however, found in word-initial open syllables in several other items from Iranic of C&B (e.g., 48. 'winter,' 50. 'rice,' 57. 'long'), so phonological conditioning with initial palatal(-alveolar) obstruents *š* (in 'stomach'), *ǰ* (in 'liver') and *g*[15] (in 'throat'), or harmonization

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

Table 5. Lexical isoglosses distinguishing Charmahali and Bakhtiari

	Persian	Charmahali	Bakhtiari
2. 'eye'	*češm*	*češ, ...*	*tiye, ...*
3. 'nose'	*bini, damāġ*	*domāġ, ...*	*noft, neft, ...*
18. 'urine'	*edrār, šāš*	*šāš, pišāb, ...*	*meste, mehse, ...*
21. 'man'	*mard*	*mard*	*piyā, mire, ...*
22. 'husband'	*šavhar*	*šuvar, ...*	*mire, ...*
23. 'wife'	*zan*	*zan, zayfe, ...*	*zine, ...*
27. 'son'	*pesar*	*pesar, ...*	*kor, ...*
31. 'bride'	*arus*	*arus, ...*	*behig, (āris,) ...*
37. 'scorpion'	*aġrab*	*aġrab, ...*	*gaždin, gādim, ...*
40. 'branch'	*šāxe*	*šāxe, ...*	*lešk, lešxa, (šāxe,) ...*
63. 's/he fell'	*oftād*	*oftād*	*vast, vahs, ...*
77. 'yesterday'	*diruz*	*diriz, diruz, ...*	*duš, ...*

with a following *i* (in 'throat' and 'white'), appears to be correlated to the retention of a high vowel.

Historical softening of b in codas. A third feature characteristic of all Iranic varieties in the C&B data is the historical softening of *b* in coda position, usually to a glide *v* ([w] in coda position), but in a couple of cases also further shifted to *y* [j] (see 4.6), coalesced with the preceding vowel, or dropped completely. This takes place in 43. 'sun' (P. *xoršid*, *āftāb*), which appears in the Iranic varieties of the province as *aftav*, *āftav*, *aftov*, *aftay*, *oftav*, *oftov* and *ofto*; 45. 'night' (P. *šab*), which is found in cognates throughout with a sonorant coda, as *šav*, *šay*, *šey*, *šev*, *šov*, *šö*, and *še*; and 47. 'water' (P. *āb*), given variously as *av*, *ov*, *o*, and *ay*.

4.4 Isoglosses between Bakhtiari and Charmahali

There is a strong bundling of isoglosses dividing Bakhtiari and Charmahali. This pattern, which includes a number of lexical items as well as several phonological correspondences, is stronger than the areal grouping of B. with Ch. against Persian (4.3). However, as signalled in the lexicostatistic analysis above (4.1) and confirmed in the table of shared correspondences later in the paper (4.7, Table 10), some of the varieties—particularly those of Boldāji (B.), Shalamzār (B.), and Shurāb Kabir (Ch.)—do not fall neatly to one side or the other. This topic is examined throughout this section. Here, we have selected the clearest isogloss patterns in the data, both lexical and phonological, but we note that there are many other patterns, not discussed in detail here, which are geographically ambiguous or indeterminate.

Distinctive Bakhtiari lexical items. In the lexicon, there are a number of widespread and distinctive B. words, as shown in Table 5.

For all of these words, and in most locations, Ch. patterns with Persian, and the B. items are distinct.

Regarding exceptions on the B. side: since Persian exhibits a marked influence on all other languages of Iran, it is not surprising when items characteristic of Persian (like the occasional B. *āris* 'bride' and *šāxe* 'branch') are used in B. However, this occurs variably between the locations where people view themselves as speakers of B.: 8 of the 12 words in Boldāji are aligned with Persian and Ch.; this is the case for 6 of the 12 words in Shalamzār; 4 in Juneqān; 2 in Fārsān and Ardal; and in two other locations, 1 of the 12 words ('bride' in both cases). At very least, this signals

Table 6. Weakening of intervocalic *b* in Bakhtiari

	Persian	Charmahali	Bakhtiari
6. 'tongue'	*zabān*	*zebun, zobun, ...*	*zevun, zovn, ...*[16]
35. 'fox'	*rubā*	*rubā, ...*	*ruvā, ruvah, ...*
79. 'tomorrow'[17]	(*fardā*)	*sobā, ...*	*sovah, sovā, ...*
80. 'day after tomorrow'	(*pasfardā*)	*passobā, ...*	*passovā, pahsovah, ...*

Table 7. Correspondences between *x* and *h*

	Persian	Charmahali	Bakhtiari
28. 'girl'	*doxtar*	*doxtar, doxdar, ...*	*dodar, dovar, ...*
49. 'house'	*xāne*	*xune, ...*	*hune, ...*
55. 'bitter'	*talx*	*talx, ...*	*tahl, ...*
58. 'dry'	*xošk*	*xošk, ...*	*hošk, ...*
60. 'red'	*sorx*	*sorx, ...*	*sohr, ...*
71. 'it burned (intr.)'	*suxt*	*suxt, soxt, ...*	*soh, sohd, ...*

overlap in the distribution of lexical structures between B. and Ch.; but it also shows the importance of structural comparisons among all varieties considered by their speakers to be "Bakhtiari," since some might be more prototypically B. than others.

The appearance of B.-type words rather than typical P./Ch. words in varieties considered by speakers to be "Charmahali" is perhaps less expected, but when it occurs, it likewise raises questions about the affiliation of these varieties. This situation does in fact arise in a few locations: Shurāb Kabir, where 3 of the 12 items in this list align with the distinctive B. form; Borujen and Arjenak, 2 items; and in Sheykh Shabān and Hāruni, each exhibiting 1 typical B. word.

For one additional item, 59. 'big' (P. *bozorg*), B. typically uses the term *gahp* or *gapu*, but the main Ch. term is *gonde*. In this case, there is a three-way isogloss distinction between Persian, Ch., and B. Exceptions for this word are found in locations overlapping those of the other lexical exceptions just mentioned: the typical Ch. form for 'big' is found in the "Bakhtiari" communities of Boldāji and Shalamzār, and the usual B. form shows up in the "Charmahali" town of Shurāb Kabir, as well as Hafshejān. The Persian word *bozorg*, perhaps a borrowing, is used alongside the regionally distinctive words for 'big' in 6 Ch. locations.

The same pattern of distinction between B. and Ch. shows up in the alignment of two recurrent phonological isoglosses in cognates: historically softened intervocalic *b* in B., and B. *h* where Ch. and P. have *x*. There are several further sound correspondences that distinguish B. from Ch., but which show up in only one or two words in the data.

Historical softening of intervocalic b. In 4.3 above, we showed that historical *b* is weakened, in codas, in all Ch. as well as B. varieties in the province. In B., this sound change typically extends to intervocalic position, as the following words show (Table 6).

In Boldāji, one of the "Bakhtiari" locations that sometimes aligns with Ch. for differentiated lexical items (as discussed in this section above and below), weakening of *b* to *v* does not apply to any

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

Table 8. Exceptions to weakening of *x* in Charmahali and Bakhtiari

	Persian	Charmahali	Bakhtiari
16. 'bone'	*ostoxān*	*ostoxun, ...*	*ostoxun, hast, ...*
17. 'blood'	*xun*	*xun, xin, ...*	*xin, (hin) ...*
64. 's/he slept'	*xābid*	*xābid, ...*	*xavsid, ...*
65. 's/he ate'	*xord*	*xord*	*xard, ...*

of the words here; but it occurs consistently in the other locations. Conversely, it shows up for 2 of the 4 words in Sheykh Shabān, a Ch. location that occasionally exhibits typical B. structures, as mentioned elsewhere in this section.

Correspondences between x and h. There are 5 items in the data where Ch. and P. *x* correspond to B. *h*. In one further item (28. 'daughter'), *x* has no corresponding phoneme in the B. data from C&B, although it is reflected as *h* in Bakhtiari varieties elsewhere (Anonby & Asadi, 2014:168). These correspondences occur in a variety of word positions, as Table 7 shows.

For several of the B. items (28. 'daughter,' 49. 'house,' 55. 'bitter,' 60. 'red'), the occurrence of *h* can be confidently traced to historical debuccalization (that is, loss of an oral place of articulation) of *x*, since Old Iranic (=OIr) sources contain this phoneme (cf. Av. *suxra* 'red') or it is consistently attested in Middle Iranic (=MIr) (Manichaean Middle Persian = MMP *duxt*, Pahlavi Middle Persian = PMP *duxtar* 'daughter'; MMP, PMP *xānag* 'house'; MMP, PMP *taxl* 'bitter').[18] However, in the case of 58. 'dry,' OIr contained an *h* (Av. *huška-*) that was subsequently hardened (before *w* and *u*) to *x* in some MIr varieties (PMP *xušk*, but cf. MMP *hušk*), so it is possible that the B. form of this word either represents a continuation of OIr *h* or, in keeping with the other words here, a late debuccalization of a hardened MIr *x*.

In 7 of the 11 B. locations, this *x*/*h* correspondence applies in all 6 items in Table 7. However, in Shalamzār and Boldāji—the "Bakhtiari" locations that pattern with Ch. for several distinctive lexical items (as discussed immediately above)—a cognate with *x* is used for all 6 words. (The segment *x* is also found in a couple of words in the southernmost locations of Lordegān and Chilteh Doderā; see 4.6.) Conversely, this typically B. sound change shows up in one "Charmahali" location—that of Shurāb Kabir—for 3 of the 6 words. This also confirms the intermediate status of Shurāb Kabir that was noted for the lexicon.

Notably, weakening of *x* to *h* does not take place anywhere in the data, including B., in three Iranic words where *x* comes from a prior historical x^w (cf. Early New Persian *ustux^{w}ān* 'bone,' PMP *x^{w}āb* 'sleep (n.),' PMP *x^{w}ard* 's/he ate'); and in a further item ('blood'), the occurrence of *h* is geographically restricted to the three north-western B. locations (Table 8; see also 4.6).

The non-application of the *x* > *h* sound change in these items suggests that it may have applied only to historical *x*, and not x^w, at a time when both phonemes were still contrastive in these varieties. In at least one B.-speaking location in the south part of the province (Milās, near Lordegān; not part of this study), the segment x^w is still used, as evidenced in the word *x^{w}ard* 's/he ate' (Mortaza Taheri-Ardali, field notes 2017). As is the case for 58. 'dry' above, the presence of *h* in 17. 'blood' in the 3 northernmost B. locations can technically be traced to a historical *h* (cf. Av. *vohuni*),[19] but a late debuccalization of a hardened MIr *x* (cf. MMP, PMP *xōn*) is also possible.

Other distinguishing sound changes and correspondences. Along with these two recurrent and generally regular B. phonological innovations, a substantial set of other sound changes and correspondences in the data distinguishes B. from Ch., although they are found in individual items rather than a large set of words. Some of these isoglosses distinguish B. from Ch. more sharply than others.

- 11. 'stomach (belly)' is found as *kom* in 9 of the 11 B. locations, and *eškam*/*ešgam* in the remaining 2 (see 61. 'white' in this list below for a similar pattern). However, in all of the 9 Ch. sites where a cognate word is used, the form *šikam* is found (cf. P. *šekam*).
- As in P. (*bače*, *bačče*), 25. 'child' has a consistent first vowel *a* in B. *bače*, etc., but *e* (Ch. *beče*) in 9 of the 11 Ch. locations. The opposite pattern is true for cognate forms of 21. 'man' (B. *merd*, Ch. *mard*) and the light verb 's/he did' (B. *kerd*, but Ch. mostly *kard*) in 67. 's/he thought.'
- With a single exception, 30. 'groom' ends with *d* in Ch. *dumād* (as in P.), but the final *d* is absent in all B. locations (*duma*, *dovā*). Similarly, Ch. *emruz*, *amruz* 'today' (cf. P. *emruz*) invariably ends with *z*, but for 7 of 11 B. locations word-final *z* is absent from this item (*amru*, *emru*, *omru*).
- In 39. 'wood' (P. *čub*), B. is found as *ču* (in line with the *b*-weakening pattern in codas given in 4.3 above), but the typical Ch. form is *čuġ*.[20] In the "Charmahali" location of Shurāb Kabir, the typical B. form *ču* is used; and the typical P. form *čub*, which is likely a recent borrowing from P., is found alongside the regional forms in four locations (3 Ch., 1 B.).
- 46. 'star' has a consistent CVC word onset in Ch. (*setāre*, = P.), but in B. a word-initial VCC sequence is prevalent (*āstāre*, *āsāre*, *ostāra*). The CVC-initial form *setāre* is used in 5 of the B. locations—alongside a typical B. form in 4 sites, and exclusively in one place. Similarly, 61. 'white' is given with a CVC word onset (*sifid*, *sefid*) in all but one of the Ch. locations (Sheykh Shaban *ispid*; also, a second form *espit* is used in Arjenak). In B., word-initial VCC sequences dominate (*espid*, *espi*, *espir*), but CVC forms (*sefid*, *safid*, *sebeyd*) are found in 4 of the 11 B. locations. As pointed out by Agnes Korn (pers. comm. 2020), it is worth noting that for both items, the prevalent B. shape aligns with MMP (*istārag*, *ispēd*) rather than with its PMP counterpart (*stārag*, *spēd*.)
- As in Persian, codas which historically contained *x* followed by a liquid have been metathesized in the Ch. words 55. *talx* 'bitter' (= P.; cf. MMP, PMP *taxl*) and 60. 'red' *sorx* (= P.; cf. PMP *suxr*, MMP *suhr*). The corresponding B. words (cf. Table 7 above) are *tahl* and *sohr*. Lack of metathesis in B. corresponds almost exactly to the historical weakening of *x* to *h*: the two "Bakhtiari" locations where metathesis occurs—Shalamzār and Boldāji—are the same places where, as in Ch., *x* has not been weakened to *h* (see the discussion of *x* > *h* in the preceding paragraphs).
- With a couple of exceptions, 64. 's/he slept' is found as *xābid* in Ch. (= P.), but as *xavsi*(*d*) in B.
- The word 65. 's/he ate' is *xord* in Ch. (= P. again in this case), but *xard* in B. (with the exceptions of Boldāji and Shalamzār, which once more follow typical Ch. forms). This distinction actually reflects two separate innovations, since the initial portions of both present-day forms originate in a historical x^wa segment (cf. PMP *x^{w}ard*; see also the discussion of *x* > *h* in this section above).

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

- 70. 'hit' contains the segment *ey* in B. (*zeyd*, *zey*), but *a* in Ch. (*zad* = P.). In Shalamzār and Fārsān (both "Bakhtiari"), the Ch./P. form is used, and in Hāruni (Ch.) both forms are attested.
- In 72. 's/he swept,' the vowel *o* appears in all B. locations (*roft*, *roh*), but in only one of the 7 Ch. locations where a cognate term is used (Sheykh Shabān); the usual Ch. form for this item contains the vowel *u* (*ruft*, *ruf*). A similar division is found, albeit less neatly, for 71. 'it burned (intr.)': the vowel *o* is used in all B. sites (*sohd*, *soxt*), and *u* in a majority of Ch. sites (*suxt*, *sux*; cf. P. *suxt*). However, for this word the 4 remaining Ch. sites employ the more typically B. form with *o*.
- To continue with 71. 'it burned (intr.)': a final *t* is found in 9 of the 11 Ch. locations (*suxt*), but only 4 of the 11 B. locations: Shalamzār, Boldāji, and the two southern sites of Lordegān and Chilteh Duderā. Remaining B. locations exhibit the forms *sohd* and *soh*.
- The words 73. 'here' and 74. 'there' contain *č* in most B. locations (*ičo*/*očo*), but *ǰ* in Ch. (*inǰā*/*unǰā*, cf. P. *inǰā*/*ānǰā*). The B. locations of Boldāji and Shalamzār once again pattern with Ch. for this feature (*unǰo*). The second portions of the two forms usually differ in both consonant voicing and vowel quality, but the phonologically intermediate forms *inǰo*/*unǰo* (also found in the Ch. locations of Hafshejān and Hāruni) suggest that they have undergone phonological changes in both directions, likely induced by language contact.

Ambivalent isoglosses. The distinction between Ch. and B. is the clearest internal division for Iranic varieties of C&B, but some isoglosses do not line up neatly with this binary distinction. The phonological isoglosses in the list immediately above are scalar, with some more clear than others. In the same way, lexical isoglosses can be ambivalent. This is the case for 42. 'leaf,' which is found as *pahr* or *par* for all of the B. locations, but also in 4 of the 11 Ch. locations, in place of or alongside the more common Ch. forms *barg* (=P.) and *balg*. Should *pahr*/*par* be considered a "typical B." item which is used in several Ch. locations, or a "typical regional Iranic" item contrasting with the "typical P." form? Debate over "typical" forms offers diminishing returns in such ambivalent cases.

Summary of distinctions between Charmahali and Bakhtiari. Taken together, the data in this section show a strong bundling of lexical isoglosses distinguishing Charmahali and Bakhtiari, and a number of sound correspondences—some recurrent in the data, and others occurring in only one or two words but with consistent and meaningful distribution between the two varieties. Other distinctive traits are ambivalent. Interestingly, the data have highlighted a number of cases where typically Ch. structures are used in locations where people view themselves as speakers of B., and vice versa. The most significant misalignments are in the "Bakhtiari" communities of Boldāji and Shalamzār and, on the other side, the "Charmahali" village of Shurāb Kabir.

4.5 Geographic variation within Charmahali

As mentioned in 3.3, differences in the language variety labels used for rural vs. urban Charmahali communities raised the possibility that there is a dialectological difference between the two groups. The preliminary lexicostatistic analysis (4.1) does not bear this out, and in fact, neither does the detailed lexical and phonological analysis here. We have reviewed each of the 80 items and found no significant structural patterns that correspond to a rural vs. urban distinction, or to any other geographic grouping of Ch. varieties.

Table 9. Shift of *m*, *n* > *v* in north-western Bakhtiari sites

	14. 'knee'	30. 'groom'	62. 's/he came'	66. 's/he knew'
Sar Āqā Seyyed	*zavi*	*dovā*	*aye*	*davest*
Deh Now Soflā	*zuvi*	*dovā*	*avod*	*dovest*
Sepidāneh	*zuni*	*dovā*	*oveyd*	*dōnest*
Bakhtiari (other)	*zuni*, ...	*dumād*	*eveyd, uma*, ...	*dunest*, ...
Charmahali	*zuni*, ...	*dumād*, ...	*umad*	*idunest*, ...
Persian	*zānu*	*dāmād*	*āmad*	*midānest*

4.6 Geographic variation within Bakhtiari

Since the Bakhtiari research locations are widely spread across the province, it is conceivable that they exhibit geographically significant patterns in the distribution of linguistic structures. In fact, this is borne out by the data. Here, we look at patterns of lexical distribution as well as phonological correspondences that we have identified in three areas of the province: the north-west corner; the south end; and B. varieties in the east and centre which, although they look like a dialectal "core" in the context of the present study, may in fact be best viewed as a periphery in relation to the wider B. language area.

The north-west corner. The clearest linguistic grouping among the B. research locations is found in the north-west sector of the province, covering three sites: Sar Āqā Seyyed, Deh Now Soflā, and Sepidāneh. Of the three locations, Sar Āqā Seyyed, located in the far north-west corner, is the most distinctive.

A couple of basic lexical items show up only in the north-west: *pange* (Sar Āqā Seyyed) and *panga* (Deh Now Soflā) are used for 9. 'hand,' in contrast to *dast* (= P.) and cognates used everywhere else; and *čel* is used only in Sar Āqā Seyyed in place of the same cognate set (*dast*) used for the meaning 8. 'arm.' Also unique to Sar Āqā Seyyed, the word *tešni* is used alongside the more common Iranic item *gelu*, *gili*, etc., for 7. 'throat' (cf. common B., common Ch., and P. *tešne* for 54. 'thirsty').

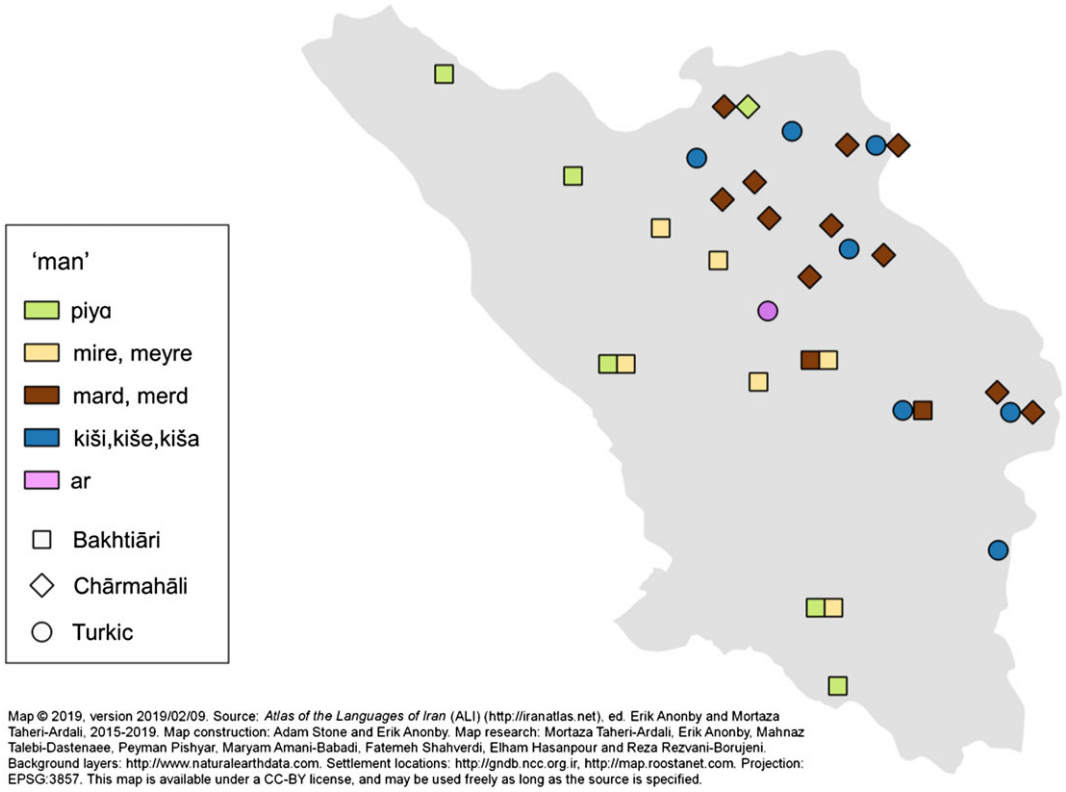
Distinctive sound changes are especially prominent in this area. As one often finds with Kurdish much further to the north (see, for example, data in Anonby, 2004/5:18), intervocalic *m* has shifted here to *v*. Interestingly, intervocalic *n* has also undergone this change in Sar Āqā Seyyed and Deh Now Soflā (Table 9).

The shift of *m* to *v* in 's/he came' (and some other words) is widespread in B. of Khuzestan Province (see Anonby & Asadi, 2014), and even in C&B it has taken place in 4 of the 8 B. locations outside the north-west: *vey* in Loshtar Gorui, *ovey* in Juneqān, *avey* in Ardal, and *eveyd* in Boldāji. Thus, for this particular item, the conservative form with *m* may be exceptional in the context of the wider B. language area.

In Sar Āqā Seyyed, the typical B. allophonic realization of *d* as [ð] following vowels and glides (Windfuhr, 1988; Anonby & Asadi, 2014:48) is taken further: *d* has merged with *r* in this position. Examples of this sound change, which takes place consistently in the data from Sar Āqā Seyyed, are as follows: 12. 'stomach' *mehre* (typical B. *mehde*); 61. 'white' *esbir* (typical B. *esbid*); and 70. 's/he hit' *zeyr* (typical B. *zeyd*).

Two other sound shifts that have taken place in Sar Āqā Seyyed are related to phonological changes already discussed. First, the weakening of *b* to *v*, which takes place in both Iranic groups (Ch. and B.) word-finally after a vowel (4.2), and also

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press



Map 6. Lexical variation in C&B Province: 'man.' From: http://iranatlas.net/module/linguistic-data.cb-lexicon-man

intervocalically in B. (4.3), extends to historical *b* as the second unit in consonant clusters in Sar Āqā Seyyed: 19. 'heart' is *qalv* (typical B. *ġalb*), and 33. 'cat' *gorve* (typical B. *gorbe*).

Second, secondary *v*—generated by the historical post-vocalic weakening of *b* to *v* (see the previous paragraph and 4.3-4.4 above)—appears to undergo a subsequent shift to *y* in Sar Āqā Seyyed: 43. 'sun' *aftay* (typical B. *aftav*); 45. 'night' *šay* (typical B. *šav*); 47. 'water' *ay* (typical B. *av*); and 62. 's/he came' *aye* (typical B. *oveyd, umad*; see Table 9 above). For 'night' in particular, it should be noted that there are also 4 Ch. locations, far away in the east part of the province, with the form *šay*, but this pattern does not show up in other Ch. words.[21] Softening of *g* to *y* in the item 32. 'dog' (*sag*, *say*) is limited to Sar Āqā Seyyed and Deh Now Soflā. However, this process occurs more widely in other words (see "Core and periphery" below).

The south end. Like the B. research sites in the north-west corner of the province, those at the south end—most clearly the southernmost site of Chilteh Duderā but also the southern city of Lordegān to some degree—exhibit distinctive structures. Distinctive lexical items are: 1. 'hair (of head)' *mel*[22] (typical B. *mi*) and 5. 'mouth' *čil* (typical B. *dohun*) in both locations; 32. 'dog' as *ketu* and *kotu* in Chilteh Duderā and Lordegān respectively (typical B. *sag*); and 73., 74., 'here,' 'there' are *iro*, *uro* (typical B. *ičo, učo*; this item is further discussed in 4.4 and 4.8). In Chilteh Duderā specifically, *fir* is used for 3. 'nose,', and the words *bot* and *xor* for 7. 'throat' (typical B. *gili*); the metathesized form *geǰar* is used for 20. 'liver' (typical B. *ǰigar, ǰiyar*); and *ġalāk* and *galāk* are both used for 41. 'stick' (typical B. *ču*, *tarke*, *gorz*), as in B. of Masjed Soleyman in Khuzestan Province (Anonby & Asadi, 2014:203). In the data, the word *dindarakul* is found for 37. 'scorpion' (typical B. *gaždin*, *gādim*) only in Lordegān. Interestingly, the words *mel* 'hair' and *bot* 'throat' are also found in the Southern Lori language area immediately to the south (Anonby, 2003b:186).

The word-final support vowel, a low vowel *a* in Middle Persian but *e* in modern Standard Persian, is phonetically somewhat unstable among the Iranic varieties of this province, but in almost all of them *e* is dominant. In Chilteh Duderā, however, it is consistently *a*. To cite some of the examples in the data: 18. 'urine' is *mehsa* (typical B. *mehse*, *mesta*); 22. 'husband' is *mira* (typical B. *mire*); 23. 'woman' is *zine* (typical B. *zina*); 25. 'child' is *bača* (typical B. *bače*); and 54. thirsty' is *tešna* (typical B. *tešne*). For this feature, the dialect of Chilteh Duderā again patterns with the word-final *a* characteristic of Southern Lori (Anonby, 2003a:92).

Core and periphery. For several lexical items, there appears to be a core/periphery division, where research sites in the north-west, west, and south share one form, and the locations in the east and centre share another. These are as follows:

- 10. 'finger' is found as *kelek* in Sar Āqa Seyyed and Deh Now Soflā (in the north-west), as well as *kilič/čelik* in Lordegān and Chilteh Duderā (in the south), contrasting with the typical B. item *angost* or *angošt* (the latter also Ch. and P.).
- 16. 'bone' is found as *hast* in Deh Now Soflā (north-west), *xas* in Loshtar Gorui (west), and *hahs* in Chilteh Duderā (south), contrasting with typical B. (as well as Ch. and colloquial P.) *ostoxun*.
- 21. 'man' is *piyā* in Sar Āqa Seyyed and Deh Now Soflā (north-west), Loshtar Gorui (west), and Lordegān and Chilteh Duderā (south), in contrast with the typical B. form *mire*. This pattern is mapped out in Map 6.

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

- 31. 'bride' is *behüg* in Sar Āqa Seyyed and *behig* in Deh Now Soflā (both in the north-west); *behig* in Loshtar Gorui (west) and Juneqān (centre-west), *beheyg* in Ardal (centre-west); *beyig* in Lordegān and *bavig* in Chilteh Doderā (both in the south). This stands apart from the B. forms common in the middle of the province (*arus*, *āris*) (as in Ch. and P.), and also used as a synonym in Loshtar Gorui, Juneqān, Ardal, and Lordegān, which originate from the Arabic equivalent *ʕarūs*.
- 33. 'cat' is found as *gulu* and *gulubiš* in Loshtar Gorui (west) and Lordegān (south), and as *gelu* (not 'throat'!—see this item in this section above) in Chilteh Doderā (also in the south), in contrast to typical B. (as well as Ch. and P.) *gorbe*.
- 51. 'egg' is found as *hāga* in Deh Now Soflā (north), both *xāve* and *xāye* in Loshtar Gorui (west), and *xāg* in Lordegān (south) (cf. MP *xāyag*; cf. also coll. P., where the meaning of the reflex *xāye* has shifted to 'testicle'[23]). In the other B. sites, the forms *toxmorġ* and *toxmemorġ* (= Ch.; cf. P. *toxmemorġ*) is used.

Rather than indicating close historical relation among the B. varieties around the periphery, the simplest explanation for these forms is that the varieties in the centre of the province have innovated. In the lexicon, this has happened through borrowing: demonstrably in the case of 'bride,' for which the central varieties use an Arabic-derived term (likely via Persian); but ostensibly also for other terms ('finger,' 'bone,' 'cat,' and 'egg'), and from Persian, since the geographically central forms of these words are similar or identical to Persian. In the case of 'man,' the B. term for 22. 'husband' *mire* has been generalized to 'man' in the same area—possibly a local semantic innovation, although equivalent shifts have been attested elsewhere in Iranic (Hassandoust, 2011:490–91). In addition, at least 4 of the 6 "peripheral" variants here are found in B. of Masjed Soleymān, at the far edge of the B. language area in Khuzestan Province to the west (see the lexicon in Anonby & Asadi, 2014). Considering all of these factors, the "peripheral" forms in C&B should most likely be viewed as shared retentions rather than shared innovations.

The same pattern holds true for phonological correspondences. 19. 'heart (organ)' retains a voiceless initial uvular obstruent (*qalb*) in Sar Āqa Seyyed (north-west) and Chilteh Duderā (south) (in keeping with B. of Masjed Soleymān), but as in Ch. and P., is voiced (*ġalb*) in all other B. locations. In the case of 20. 'liver' (P. *jegar*), historical *g* has been softened to *y* (*ǰiyar*) in the 3 north-western sites (Sar Āqā Seyyed, Deh Now Soflā, and Sepidāneh) as well as Loshtar Gorui (west), Juneqān, and Ardal (centre-west). In the remaining 5 B. sites, which are found in the centre and south areas of the province, the *g* characteristic of the Ch. form *ǰigar* (and P. *ǰegar*) is retained. Similarly, for 10. 'finger' in the 6 remaining locations (cf. *kelek* etc. in the list immediately above), the first consonant of the word-final cluster is *s* (*angost*) in the more peripheral locations of Deh Now Soflā, Sepidāneh, and Ardal, and *š* (*angošt*, = Ch., P.) in the 3 locations toward the centre of the province (Boldāji, Shalamzār, and Fārsān). Both variants are attested in Juneqān (centre-west). The same pattern shows up for 53. 'hungry,' generally *gosne* in B., but *gošne* in the same 4 central locations. This brings us full circle to the isoglosses distinguishing Ch. and B. (4.4), where the "Bakhtiari" locations of Boldāji and Shalamzār in particular often pattern with Ch.

4.7 Tabulation of sound correspondences between Iranic varieties

In the preceding sections (4.2–4.6), we delineated 80 instances of sound correspondences within cognate sets that show clear geographic patterning. In addition, we have identified 34 other instances of recurrent but (what appear to us to be) geographically ambiguous sound correspondences. The complete list of 114 correspondences is inventoried in Appendix 2.

Using the program Wordsurv, we have grouped and tabulated individual correspondences for 24 Iranic language varieties: the 22 Iranic varieties in our sample from C&B, along with Standard Persian and Bakhtiari of Masjed Soleymān as reference varieties. In the following table (Table 10), as for the table of cognate comparisons in 4.1 above (Table 3), B. of Masjed Soleymān (MJS, top row) is followed by 11 other B. varieties from Sar Āqā Seyyed (SAS) in the north-west to Chilteh Duderā (CHT) in the south. Ch. varieties follow a similar progression across the north-east corner of the province from Cham Chang (CHC) in the north to Naqneh (NQC) in the south-east. Persian (PRS) is shown in the bottom row of the table. Percentage of shared values for these correspondences is as follows (Table 10).

These percentages of shared sound correspondences provide a picture of the language situation which is finer-grained than the results of lexicostatistic analysis (4.1) but confirms the same overall patterns. It also helps to visualize trends in the subsequent analysis sections above (4.2–4.6). When focusing on sound correspondences, Persian (bottom line) is recognizably distinct from both Ch. and B., but much more so from B. (cf. 4.3). B. and Ch. pattern differently for many of the sound correspondences although, as highlighted in 4.4 above, forms given by B. speakers from Shalamzār (SHL) and Boldāji (BOL) in particular pattern with Ch. varieties more closely than with some B. varieties. Conversely, the Ch. variety of Sheykh Shabān (SHS) shows a high percentage of shared sound correspondences with most B. varieties. In keeping with the observations given in 4.5, Ch. is an internally coherent grouping. B., however, covers a large geographic area across the province and is internally heterogeneous. As outlined in 4.6, the north-west corner, the south end, and the entire western periphery of the B. language area each show higher levels of internal similarity than they do with the B. language area as a whole. The B. geographic "core," which includes Fārsān (FRS), Juneqān (JNB), Shalamzār (SHL), and Boldāji (BOL), shows some internal consistency as well as shared similarity to neighbouring Ch. varieties.

4.8 Turkic and Iranic

Up to this point, we have focused on analysis of the Iranic (Ir.) varieties of C&B (4.2–4.6), both because they cover most of the province and because the internal relations between these varieties are key to understanding the language situation. However, there are also important points to be elaborated regarding the internal dialect structure of the province's Turkic (T.) varieties, as well as structural convergence between T. and Ir. varieties in the region. Within the scope of the present paper, it is not possible to give a complete account of these topics, but a fuller analysis is undertaken elsewhere (Schreiber et al., 2017, 2021; Anonby, Schreiber & Taheri-Ardali, 2020; Anonby, Taheri-Ardali, Schreiber et al., 2020).

Linguistically, the 8 T.-speaking locations where data have been collected are relatively homogeneous. The differences between them are generally due to varying degrees of convergence with neighbouring Ir. varieties—both B. and Ch., depending on the part of the province. This topic is introduced briefly in this section, but is treated in more detail in Schreiber et al. (2017, 2021).

As expected, and as shown already in the preliminary lexicostatistic analysis (4.1), the T. varieties of the province are

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

Table 10. Percentage of shared values for identified sound correspondences

	MJS	SAS	DNC	SEP	LGI	FRS	JNB	ARD	SHL	BLB	LDG	CHT	CHC	SKC	SHS	ARJ	FTA	HAR	SKO	FSH	HAF	BOR	NQC	PRS
MJS	100	77	83	88	79	73	78	79	63	57	70	61	44	39	60	48	54	49	41	40	44	41	42	31
SAS	77	100	79	75	64	69	70	73	55	48	63	54	41	35	55	40	44	40	31	30	35	32	33	29
DNO	83	79	100	88	72	73	80	83	67	57	71	66	46	42	62	49	55	51	39	41	46	45	40	32
SEP	88	75	88	100	80	77	86	88	71	64	78	71	52	47	66	54	61	57	46	45	50	46	43	39
LGI	79	64	72	80	100	66	78	81	61	60	80	77	51	47	61	53	58	54	49	48	51	47	48	33
FRS	73	69	73	77	66	100	84	77	74	71	69	62	59	57	74	64	72	65	56	56	61	58	54	41
JNB	78	70	80	86	78	84	100	87	75	75	74	69	60	56	71	63	71	66	55	53	59	56	52	44
ARD	79	73	83	88	81	77	87	100	67	68	79	70	56	53	70	58	66	61	50	50	56	51	48	42
SHL	63	55	67	71	61	74	75	67	100	82	69	59	74	68	81	78	78	80	70	65	75	68	65	57
BLB	57	48	57	64	60	71	75	68	82	100	67	59	78	76	75	79	80	84	77	70	78	74	71	61
LDG	70	63	71	78	80	69	74	79	69	67	100	85	59	53	64	62	64	60	58	53	58	54	53	43
CHT	61	54	66	71	77	62	69	70	59	59	85	100	47	42	56	49	53	49	45	41	49	43	42	27
CHC	44	41	46	52	51	59	60	56	74	78	59	47	100	95	80	93	87	94	93	89	86	94	90	78
SKC	39	35	42	47	47	57	56	53	68	76	53	42	95	100	71	93	84	91	96	89	89	96	91	75
SHS	60	55	62	66	61	74	71	70	81	75	64	56	80	71	100	77	79	75	69	65	75	73	71	54
ARJ	48	40	49	54	53	64	63	58	78	79	62	49	93	93	77	100	88	94	94	93	93	92	91	64
FTA	54	44	55	61	58	72	71	66	78	80	64	53	87	84	79	88	100	93	81	79	90	83	79	60
HAR	49	40	51	57	54	65	66	61	80	84	60	49	94	91	75	94	93	100	90	86	91	89	86	68
SKO	41	31	39	46	49	56	55	50	70	77	58	45	93	96	69	94	81	90	100	91	90	95	91	73
FSH	40	30	41	45	48	56	53	50	65	70	53	41	89	89	65	93	79	86	91	100	90	90	94	61
HAF	44	35	46	50	51	61	59	56	75	78	58	49	86	89	75	93	90	91	90	90	100	91	88	65
BOR	41	32	45	46	47	58	56	51	68	74	54	43	94	96	73	92	83	89	95	90	91	100	96	73
NQC	42	33	40	43	48	54	52	48	65	71	53	42	90	91	71	91	79	86	91	94	88	96	100	66
PRS	31	29	32	39	33	41	44	42	57	61	43	27	78	75	54	64	60	68	73	61	65	73	66	100

(place names represented here by three-letter research location codes are defined in Table 2 above)

unquestionably distinct from the Ir. varieties, and the majority of structures are unrelated. This is shown for a representative sampling of lexical items, shown in Table 11, for which there are no cognates shared between the two families.

Interestingly, for all 11 verbs in the data, there are no cases where verbal roots are borrowed from Iranic into Turkic.[24]

Still, there is significant overlap for some lexical items. For some words (20. 'liver,' 40. 'branch,' 42. 'leaf,' 56. 'fresh'), borrowing from Ir. has taken place in all of the T.-speaking sites in the data, but for other words (19. 'heart,' 35. 'fox,' 38. 'tree,' 79. 'tomorrow') it occurs only sporadically. A cursory observation from our limited data set, as evidenced by some of these items, points to nature-related vocabulary as one domain that is susceptible to borrowing.

Contact-induced change in Turkic of Iran is often attributed to influence from Persian (Kıral, 2000; Erfani, 2012), perhaps because Persian is better known than other Ir. varieties in direct contact with Turkic. Since, as we have shown, Ir. varieties of C&B differ in many respects from Standard Persian and from one another—and, as Map 1 shows, some T.-speaking communities are in contiguous to B., and others are beside Ch.—it is possible to trace the path of language contact more precisely here.

Technical vocabulary, such as medical terminology like 11. 'stomach' and 19. 'heart,' has been borrowed from Arabic into Persian and from there into most other Ir. varieties, so it is not unlikely that Persian could serve as a direct source language for these items in T. as well. For both of these items, a Persian-type word is found in 7 of the 8 T. locations (Table 12), although in one or two cases for each word, is it used alongside a T. term.

In Persian, the term 'red' is expressed as either *ġermez* or *sorx*. Whereas almost all B. and Ch. varieties in the data show cognates of *sorx* (but admit *ġermez* as an alternate term in three locations), in

Table 11. Examples of distinct Turkic vocabulary in C&B

	SW Iranic (P., Ch., B.)	Turkic
4. 'ear'	*guš*	*gulāġ, gulāx*
8. 'arm'	*dast, das*, ...	*al*
14. 'knee'	*zānu, zuni*, ...	*diz*
28. 'girl'	*doxtar, dohdar*, ...	*ġez, ġiz*, ...
34. 'wolf'	*gorg*	*ġurt*
47. 'water'	*āb, av*, ...	*su*
61. 'white'	*sefid, esbid*, ...	*āġ*
65. 's/he ate'	*xord, xard*, ...	*yede*, ...
71. 'it burned (intr.)'	*suxt, sohd*, ...	*yānde*, ...
75. 'under'	*zir*, ...	*ālt*

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

Table 12. Technical vocabulary borrowed into Turkic

	Arabic	Persian	Charmahali	Bakhtiari	C&B Turkic
11. 'stomach'	*maʕda*	*me'de*	*mehde*, ...	*mehde*, ...	*mehda*,[25] ...
16. 'heart (organ)'	*qalb*	*ġalb*	*ġalb*, ...	*ġalb*, ...	*ġalb*, ...

Table 13. Evidence of local borrowings into Turkic

	Persian	Charmahali	Bakhtiari	C&B Turkic
12. 'stomach'	*fardā*	*sobā*, ...	*sovā*, ...	*sobā, sahar, sovā*, ...
20. 'liver'	*ǰegar*	*ǰigar*	*ǰiyar*, ...	*ǰigar, ǰiyar*
35. 'fox'	*rubāh*	*rubā*, ...	*ruvā*, ...	*tilki, ruvā, rubā*
42. 'leaf'	*barg*	*barg, balg*, ...	*pahr*, ...	*barg, pahr*, ...

T. 'red' is invariably cognate with *ġermez*.[26] This suggests that the T. word for 'red' has been borrowed from Persian rather than the Ir. languages of C&B.

However, in other cases it is more clearly the local Ir. languages which are lexifiers (Table 13).

Further underlining the local character of borrowing, there are direct correspondences between the particular forms used by T. speakers in a given location, and the equivalent structures—cognate grouping as well as the specific phonological content—in the adjacent Ir. varieties. To illustrate: in the data, T. always borrows the Ir. word for 'leaf,' but even more specifically, it usually exhibits the term *pahr* when near B. areas that use *pahr*, and *barg* next to Ch. areas that use *barg*. This is shown for cognate groupings of 'leaf' in the following map (Map 7).

Several additional words provide probable evidence of contact, but the source and direction of borrowing is less clear.

First, the term *ǰeġele* shows up for 26. 'boy' in all three languages: 8 times for Ch. (usually alongside other Ir. terms), 4 times for B., and once for T. (with the variant *ǰeġela*). The conspicuously similar terms *oġol*, *oġlān*, and *oġel* are more common for T. Are all of these words related? Was the term *ǰeġele* borrowed from T. into Ir., but (given the significant yet consistent differences between the forms) at an earlier point in history?

Item 52. 'walnut' also presents a puzzle. In all of the Ir. varieties in the data sample, this word is rendered as *gerdu*, which itself is probably an innovation in Ir.; the Middle Persian term, and that of many Ir. varieties today, is *gōz*. In Arabic (outside of this study), the equivalent is *ǰawz* (and similar variants), and in all of the T. varieties in the data the word *ġoz* is given. Perhaps the T. term has been borrowed from Persian (or elsewhere in Ir.), but this must have happened at an early point in time, since the vowel *o* (attributable to the MIr vowel *ō* rather than the New Ir. vowel *ū/u*) is used.

The *Wanderwort*[27] 33. 'cat' presents a further riddle. In both Ir. groupings in the province, the geographically dominant term is *gorbe* (= P.). Two T. varieties in the data also use a cognate of *gorbe*, but the related forms *pišig*, *püšüg*, and *pišug* are used in five other T. locations (and *mali* in a final, single T. site).[28] The form *pišuli*, found in Ch. of Shahr-e Kord, clearly patterns here with the more usual T. forms. Has it been borrowed from local T.? Do all of the *piš*-type words for cat originate in an earlier (possibly) Ir. prototype, since a similar term is found in distantly-related Ir. varieties?[29] Or perhaps both explanations are relevant?

As mentioned in 4.6, the words 73., 74., 'here,' 'there' are *iro*, *uro* in Lordegān and Chilteh Duderā, the two southernmost B. sites. These terms appear to combine the phonological content of typical B. locationals *ičo/učo* with T. equivalents *bura/burā* and *ora/orā*. (Alternatively, the B. term *rah* 'way, path' (Anonby & Asadi, 2018:217; P. *rāh*), and/or historical *r*-type locationals attested elsewhere in Iranic[30] may have contributed to both Turkic and Bakhtiari *r*-type locationals.)

Further, the main term for 7. 'throat' in the T. data is *boġāz*, similar to the term *boġāzi*, used in some Southern Lori varieties to the south (Anonby, 2003b:186), and the term *xerrex*, found in two T. locations, resembles the term *xor* in the southerly B. location of Chilteh Duderā, and even more closely the B. form *xer* attested elsewhere (ibid.). All of these lexicon-related scenarios deserve further analysis in the light of additional comparative material from T. and Ir.

Two final cases of structural convergence between T. and Ir. come from the phonology. The first concerns front rounded vowels *ü* and *ö*. Although front rounded vowels are not a typical feature of Southwestern Ir., they appear sporadically across Iranic (e.g., Okati et al., 2010), including a number of the Ir. varieties in C&B that are in contact with T. (see also 4.3 above). Conversely, as occurs intermittently in T. varieties elsewhere in Iran (especially Qashqai of Fars Province; see Bulut, 2016), front rounded vowels have been lost in the phonological inventories of some of the T. varieties in C&B, presumably as a consequence of contact with neighbouring Ir. varieties. Map 8 shows that the presence vs. absence of front rounded vowels in C&B is more easily attributable to areal similarity than to the family—T. or Ir.—to which a variety belongs.[31]

A second sound correspondence which also appears to be shared is shift of postvocalic *v* to *y*. For Iranic, this occurs in three words in Sar Āqā Seyyed (B.; see 4.8), for which a shifted form for 48. 'night' is shared with 4 Charmahali locations. This occurs separately for 4 of the 8 Turkic locations in 49. 'house,' which is evenly divided between the more typical Turkic *ev* (also *öv* in one location) and the local reflexes *ey*, *öy*.

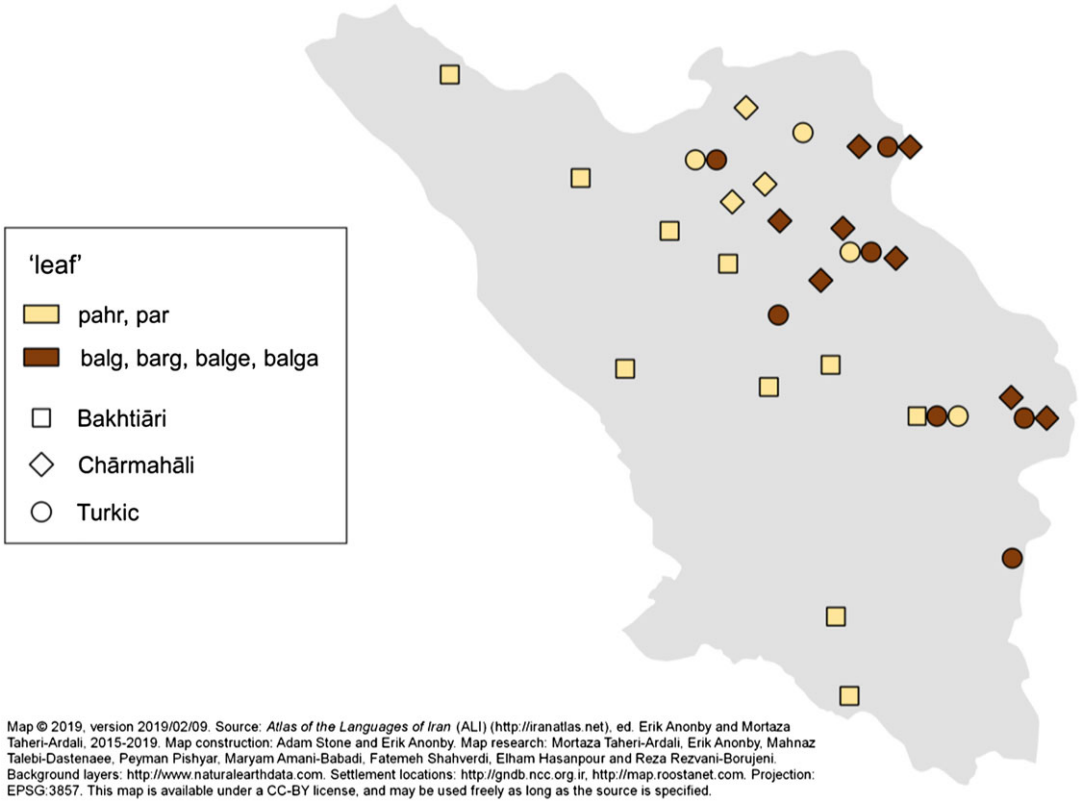
5. Findings and prospects

Over the history of documentation of the languages of Iran, going back more than a century, Chahar Mahal va Bakhtiari (C&B) Province has been for the most part passed over. While a few studies have appeared on Bakhtiari varieties of the province, its Charmahali and Turkic varieties are almost absent in the literature. This paper seeks to provide a picture of the language situation in C&B which is both global and fine-grained, although still initial.

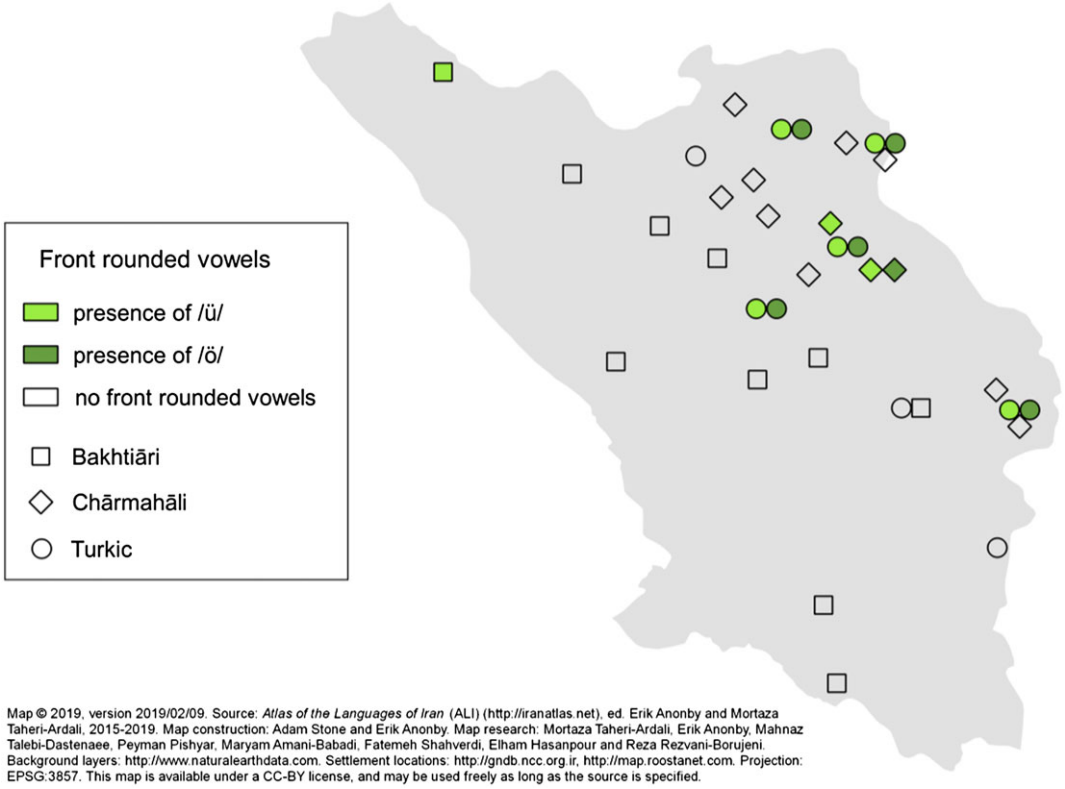
Our study was conducted in the context of the *Atlas of the Languages of Iran* (ALI) research programme (section 2). It opens with a description of the activities that were necessary to frame and enable a coherent understanding of C&B province as a linguistic area, including an initial period of literature review (3.1) and a longer phase of language distribution research (3.2). Reflection on the detailed, emergent picture of language distribution (3.3) raised key research questions and facilitated the selection of sites (3.4) for linguistic data collection with the ALI questionnaire (3.5).

The second half of the paper (section 4) treats the linguistic data collected by the research team in 30 sites across the province between 2015 and 2018: 11 Bakhtiari, 11 Charmahali, and 8 Turkic. Because of the richness of the lexical data set—80 words

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press



Map 7. Lexical variation in C&B Province: 'leaf.' From: http://iranatlas.net/module/linguistic-data.cb-lexicon-leaf



Map 8. Phonological variation in C&B Province: Front rounded vowels. From: http://iranatlas.net/module/linguistic-data.cb-phonology

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

in each of the 30 varieties—in this first study we limit our analysis to lexicon, along with phonological correspondences associated with cognate sets. In our general overview of the province, we look at all three language groupings. However, our detailed analysis focuses on relationships and internal structure of the Iranic varieties. The discussion of Turkic is cursory here, concentrating on issues of language contact; analysis of internal relationships among the Turkic varieties is reserved for a separate study.

Even within the limitations of the scope of this study, a first clear overview of the language situation emerges through analysis of the lexical and phonological data, and subsequent reinterpretation of language distribution data in light of this analysis. Preparatory lexicostatistic analysis (4.1) makes it clear that the lexicon of Turkic in C&B is fundamentally different from that of Iranic, and it also shows structural divergences between Bakhtiari, Charmahali, and Persian. However, it does not confirm any structural distinction between the Rural and Urban Charmahali groupings suggested by the labels that speakers use.

The more detailed analysis we have carried out shows a number of significant patterns, providing insight into the questions raised in 3.3 above.

1) Charmahali and Bakhtiari exhibit a strong base of lexical items and phonological forms shared with Persian, confirming their place alongside Persian within the Southwestern branch of Iranic (4.2).
2) There is a small but clear set of lexical items and shared sound changes common to the Iranic varieties of the region (Ch. and B.) but distinct from Persian (4.3).
3) A robust bundling of isoglosses distinguishes Charmahali and Bakhtiari, both for lexicon (where 12 of 80 items show a general binary distinction) and for phonological patterning in cognates, where about a dozen diagnostic correspondences of varying strength have been identified (4.4). In most of these cases, the Charmahali structures are aligned with Persian, and sometimes even identical to it. In a small number of cases, each of the three groupings (P., Ch., B.) exhibits a distinctive form. With the exception of a single item (25. 'child'), alignment of Bakhtiari lexicon with Persian, to the exclusion of Charmahali, is not attested in the data. A binary distinction between Charmahali and Bakhtiari does not pattern neatly for all locations, however. Bakhtiari varieties in the centre of the province, and in particular Boldāji and Shalamzār, show a higher level of lexical and phonological similarity with neighbouring Charmahali varieties than they do with Bakhtiari varieties at the geographical peripheries of the province (see point 6 below). On the other side, in the Charmahali village of Shurāb Kabir, many structures are shared with Bakhtiari. Other exceptions to the general distinction between Charmahali and Bakhtiari can be attributed to widespread adoption of Persian forms for specific words.
4) Although we considered a possible Rural vs. Urban Charmahali dialect distinction for the 80 words in the data, were we unable to identify any structural patterns that distinguished the two groups, or any other dialectological group within Charmahali (4.5).
5) Within Bakhtiari, the geographic distribution of lexical and sound correspondence patterns support the positing of a north-west dialect area; a separate dialect zone in the south; and a surprisingly well-defined core/periphery pattern, where Bakhtiari dialects in the centre of the province exhibit innovations that distinguish them from dialects in the north-west, west, and south (4.6). What first appears to be a dialectal "core" area in the centre of the province, is likely better viewed as a periphery in the context of the Bakhtiari language area as a whole. We have confirmed the divergence of Bakhtiari in the central area by comparison with lexicon and sound correspondences in the Bakhtiari reference variety of Masjed Soleymān in Khuzestan Province. The general pattern of divergence in the central area, which includes the two strongly Charmahali-leaning Bakhtiari locations of Boldāji and Shalamzār, appears to be due to areal convergence between Charmahali and Bakhtiari or the influence of Persian.
6) The Turkic varieties of C&B are fundamentally different from the Iranic varieties (Ch. and B.), but there are significant patterns of contact-induced change. Most cases are examples of lexical borrowing of Iranic vocabulary into Turkic (4.8). However, the source and path of contact effects are uncertain or, in fact, ambivalent—pointing in both directions—for several lexical items, as well as two phonological features: the shared geographical distribution of front rounded vowels *ü* and *ö* in the data, and an apparent shared *v* > *y* sound change.

Taken together, these findings offer a first fine-grained and geographically representative overview of the language situation in Chahar Mahal va Bakhtiari Province, and have helped to set the direction of the research process for other provinces in the *Atlas of the Languages of Iran*, including the refinement of the ALI linguistic data questionnaire. They also point to outstanding questions for ongoing work on C&B Province: Will morphosyntactic data confirm the emergent picture of the language situation, or will they pattern differently from the lexical and phonological data? What are internal relationships among Turkic varieties of the province, and how do these varieties fit in to Turkic of Iran more generally? As we wondered earlier but did not investigate here, does the Persian emerging in C&B have an areal character, whether as a result of contact with other languages in the province, or a substrate that shows up among speakers that have shifted from these other languages? And in the wider scheme, to what degree do the boundaries of C&B as a linguistic area—in the more technical sense of *Sprachbund*—follow the province's borders? How will an extension of our research to related language varieties in other provinces impact the way that we understand the language situation in C&B Province?

Here, we add some final notes regarding language identification, which was a further issue raised in 3.3 above, and one which necessarily shapes the way that linguists frame and discuss the language varieties of the province. Although such assessments are always to some degree subjective, it is clear to us, based on speakers' perspectives as well as structural considerations highlighted in the analysis, that Bakhtiari can be considered a language distinct from Charmahali and Persian, although there is a significant level of convergence between Bakhtiari and Charmahali in the central areas of the province. The status of Charmahali, especially in relation to Persian, is less clear even when the question is considered from both of these directions. Will a detailed analysis of morphosyntax shed further light on the depth and pervasiveness of Charmahali's structural distinctiveness, or will it continue to yield ambivalent results? Or, perhaps, comparison with neighbouring Southwestern Iranic varieties in Esfahan Province will bring clarity to the issue. Finally, for communities of speakers that consider themselves "Bakhtiari" or "Charmahali" but use structures more

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

typical of the other language group, studies of language contact situations and oral histories of migration can be an important means of untangling the threads of the province's linguistic tapestry.

Acknowledgments. Earlier versions of portions of this paper were presented by Erik Anonby and Mortaza Taheri-Ardali at the 7th International Conference on Iranian Linguistics (ICIL7), Lomonosov Moscow State University, 28–30 August 2017, and by Erik Anonby at the Linguistics Summer School Evening Lecture Series, Leiden University, 16 July 2018. The authors wish to thank each of the 44 consultants who contributed data for this study. We also acknowledge the dedicated and valuable contributions of ALI colleagues for Chahar Mahal va Bakhtiari (C&B) Province in carrying out field research for this article; a full list is provided in section 3.1. Finally, we are grateful to Agnes Korn, Don Stilo, and an anonymous reviewer for detailed critique on specific aspects of this study. Various elements of the research were supported by SSHRC (Social Sciences and Humanities Research Council of Canada), Carleton University, Universität Bamberg, and the Alexander von Humboldt Foundation.

Data availability statement. Consultant and researcher names, other metadata, and linguistic data for each research location are available in the *ALI Dataverse* at the permanent link: https://doi.org/10.5683/SP2/FVLDLZ.

Notes

1 Abbreviations: Av.: Avestan; B.: Bakhtiari; C&B: Chahar Mahal va Bakhtiari (Province); Ch.: Charmahali; coll.: colloquial; Ir.: Iranic; MIr: Middle Iranic; MMP: Manichaean Middle Persian; PMP: Pahlavi Middle Persian; OIr: Old Iranic; P.: Persian; T.: Turkic. Historical data sources: Avestan: Peterson (1995); Pahlavi Middle Persian: MacKenzie (1971); Manichaean Middle Persian: Durkin-Meisterernst (2004).

Transcription of phonological data follows the conventions of Iranian linguistics set out in https://carleton.ca/iran/transcription, with the following phonemic symbols differing from their phonetic counterparts in IPA: *ạ* [aˤ], *č* [t͡ʃ], *ġ* [ɢ]/[ʁ], *ǰ* [d͡ʒ], *š* [ʃ], *v* [ʋ]/[v]/[w] and *x* [χ].

2 A list of institutional partners and research team members is found at: http://iranatlas.net/module/atlasteam.

3 See: http://iranatlas.net/module/bibliography.

4 See: http://iranatlas.net/module/classification.

5 Initial results from this phase were also verified and extended for each of the sites where the questionnaire was subsequently carried out across the province (3.4, 3.5).

6 The historical districts of a) Lār, b) Kiār, c) Mizdej, and d) Gandomān are now respectively located in the *šahrestān* (provincial sub-districts) of a) Shahr-e Kord, Sāmān, and Ben, b) Kiār, c) Fārsān, and d) Borujen. The history of these areas is recounted in Sāsānpur (1393/2014).

7 This term is used by people in Iran refer to any variety of Turkic, including Azerbaijani and Standard Turkish of Turkey; see Bulut (2014:19).

8 The programme of the workshop, along with a list of participants and topics, is available at: https://www.uni-bamberg.de/aspra/workshop-questionnaire-languages-of-iran-2017.

9 The colour scheme for this table, automatically generated by Wordsurv, shows cells with high percentages of lexical similarity as green and cells with low percentages as red. A full range of intermediate percentages, which pass through yellow (relatively higher similarity) and orange (relatively lower similarity) on the colour wheel, is not represented in the data.

10 There is not 100% similarity in all directions among all four locations because of cases where two or more lexical items, sometimes with different cognate values, are given for some items in the wordlist.

11 The synchronic analysis of contrastive vowel lengthening as an allophonic realization of post-vocalic *h* is discussed in Anonby & Asadi (2014:59–60). This interpretation applies to equivalent examples throughout the data.

12 Use of ellipsis (. . .) in data tables indicates that other, less frequent lexical forms are attested for some research sites.

13 An anonymous reviewer has pointed out that the verb *roft* 's/he swept' was formerly used in Persian as well, but is now obsolete. This enriches the historical picture further.

14 The latter form *periz* 'yesterday,' which is used in the Charmahali-speaking city of Hafshejān, has been verified. Although *periz* resembles the word for 'day before yesterday' in Persian (*pariruz*) and other Ch. varieties (*peyruz*, etc.), it contrasts with *peririz* 'day before yesterday' in the variety of Hafshejān.

15 Throughout C&B province, as in much of Iran, *g* has a palatal allophone [ɟ] before front vowels.

16 The word 'tongue' is *zabān* in Early New Persian as well as contemporary Standard Persian, but is also attested as *uzwān* or *zuwān* in MP. It is therefore possible that the Bakhtiari form for this word displays retention of *w* (=*v*) rather than softening of *b*; and that it is the forms with *b* that have innovated through hardening of *v*.

17 The Charmahali and Bakhtiari equivalents for 'tomorrow' and 'day after tomorrow' shown here are borrowings from Ar. *sabh* 'morning,' as explained in 4.3 above.

18 This sound shift is probably also relevant for 71. 'it burned,' which is represented in Middle Iranic by PMP *sōxt* and MMP *sōz-*, and for 51. 'egg,' found with *h* in 2 northern B. locations (*hāye* in Sar Āqā Seyyed and *hāga* in Deh Now Soflā, cf. MMP, PMP *xāyag*) but for which no cognate is found in Ch.

19 Historical phonology of the word 'blood,' which shows irregularities in its patterning across West Iranic, is treated in Schwartz (1982) and Cathcart (2015).

20 Habib Borjian (pers. comm. 2017) notes that the form *čuġ* 'wood' is also used in Persian varieties of Esfahan Province.

21 Regarding *y* in this word, see the discussion (in this section below) of a subsequent *v* to *y* shift in Sar Āqā Seyyed and a similar shift in Turkic varieties of the province (4.8).

22 In other Bakhtiari varieties, the word *mel* refers to body hair rather than hair of the head (Anonby & Asadi, 2014:158); here, it has been generalized to include all types of hair.

23 As pointed out by an anonymous reviewer.

24 For two verbs, however, the nominal component of a light verb construction has been borrowed into Turkic.

25 The use of word-final *a* as a support vowel and replacement of glottal stop with *h* are in keeping with the phonological system of C&B Turkic (Anonby et al., in preparation).

26 Juneqān, which is a majority Turkic-speaking town, is the single B. site where *ġermez* is the only form used.

27 A word that is introduced into a language along with the adoption of an associated item or activity.

28 This form is similar to the term for cat in some Central Plateau Iranic varieties: *molǰi* (Rāji of Abuzeydabad) and *maṛǰine* (Rāji of Barzok) (Talebi-Dastenaee & Anonby, forthcoming; Talebi-Dastenaee et al., under consideration; Mahnaz Talebi-Dastenaee, field notes 2018).

29 For example, Kalhuri Kurdish *pıšī* (field notes, Mojtaba Gheitasi et al., 2017; field notes, Negar Sherafat et al., 2017).

30 For historical forms in Old and Middle Ir., see Hassandoust (2011:943–46); present-day locationals with *r* are attested from languages such as Southern Kurdish (Fattāh, 2000:652–53).

31 Although it is not visible from the map, which shows only C&B, the Bakhtiari community of Sar Āqā Seyyed in the far north-west—where the front rounded *ü* is phonemic (Taheri-Ardali, 2017a)—is in contact with Turkic-speaking people across the border of Esfahan Province (Mortaza Taheri-Ardali, field notes 2017).

References

Aliakbari, Mohammad, Mojtaba Gheitasi & Erik Anonby. 2014. On language distribution in Ilam Province, Iran. *Iranian Studies 48*(6). 1–16.

Anonby, Erik. 2003a. *A phonology of Southern Luri*. Munich: Lincom Europa.

Anonby, Erik. 2003b. Update on Luri: How many languages? *Journal of the Royal Asiatic Society 13*(2). 171–97.

Anonby, Erik. 2004/5. Kurdish or Luri? Laki's disputed identity in the Luristan province of Iran. *Kurdische Studien 4/5*. 7–22.

Anonby, Erik & Ashraf Asadi. 2014. *Bakhtiari studies: Phonology, text, lexicon*. Studia Iranica Upsaliensia 24. Uppsala: Acta Universitatis Upsaliensis.

Anonby, Erik & Ashraf Asadi. 2018. *Bakhtiari studies II: Orthography*. Studia Iranica Upsaliensia 34. Uppsala: Acta Universitatis Upsaliensis.
Anonby, Erik, Mojtaba Gheitasi & Mohammad Aliakbari. 2017. Revisiting language distribution in Ilam Province. 3rd International Conference on Kurdish Studies (ICKS3), University of Exeter, June 26–28, 2017.
Anonby, Erik, Amos Hayes & Robert Oikle. 2020. A multi-dimensional approach to classification of Iran's languages. In Richard K. Larson, Sedigheh Moradi & Vida Samiian (eds.), *Advances in Iranian linguistics* (Current Issues in Linguistic Theory 351), 29–56. Amsterdam: John Benjamins.
Anonby, Erik, Masoud Mohammadirad & Jaffer Sheyholislami. 2019. Kordestan Province in the Atlas of the Languages of Iran: Research process, language distribution, and language classification. In Songül Gündoğdu, Ergin Öpengin, Geoffrey Haig & Erik Anonby (eds.), *Advances in Kurdish linguistics*, 9–38. Bamberg, Germany: Otto-Friedrich-Universität Bamberg.
Anonby, Erik & Parisa Sabethemmatabadi. 2019. Representing complementary user perspectives in a language atlas. In D.R. Fraser Taylor, Erik Anonby & Kumiko Murasugi (eds.), *Further developments in the theory and practice of cybercartography: International dimensions and language mapping*, 413–40. Amsterdam: Elsevier.
Anonby, Erik, Laurentia Schreiber, Mortaza Taheri-Ardali & Geoffrey Haig. In preparation. Turkic of Chahar Mahal va Bakhtiari Province.
Anonby, Erik, Laurentia Schreiber & Mortaza Taheri-Ardali. 2020. Balanced bilingualism: Patterns of contact influence in L1 and L2 Turkic and Bakhtiari speech in Juneqan, Iran. *Iranian Studies 53*(3/4). 589–622.
Anonby, Erik & Mortaza Taheri-Ardali. 2017. Linguistic diversity and language contact in Chahar Mahal va Bakhtiari Province, Iran. 7th International Conference on Iranian Linguistics (ICIL7), Lomonosov Moscow State University and the Institute of Linguistics of the Russian Academy of Sciences, Moscow, August 28–30, 2017.
Anonby, Erik & Mortaza Taheri-Ardali. 2018. Bakhtiari. In Geoffrey Khan & Geoffrey Haig (eds.), *The languages and linguistics of western Asia: An areal perspective*, 445–80. Berlin: De Gruyter.
Anonby, Erik, Mortaza Taheri-Ardali, et al. 2015–2022. *Atlas of the languages of Iran*. Ottawa: Geomatics and Cartographic Research Centre (GCRC), Carleton University. http://iranatlas.net (14 January 2022).
Anonby, Erik, Mortaza Taheri-Ardali, Geoffrey Haig, et al. 2020. Atlas of the Languages of Iran (ALI) questionnaire. In Erik Anonby, Mortaza Taheri-Ardali, et al. (eds.), *ALI Dataverse*. Toronto: Scholars Portal Dataverse. https://doi.org/10.5683/SP2/SDJ5N4.
Anonby, Erik, Mortaza Taheri-Ardali & Amos Hayes. 2019. The Atlas of the Languages of Iran (ALI): A research overview. *Iranian Studies 52*(1–2). 199–230. https://doi.org/10.1080/00210862.2019.1573135.
Anonby, Erik, Mortaza Taheri-Ardali, Laurentia Schreiber, Christiane Bulut, Geoffrey Haig, Peiman Pishyar Dehkordi, Mahnaz Talebi-Dastenaee, Fatemeh Shahverdi, Hossein Hashemi Zarajabad, Masoud Mohammadirad, Faranak Jamaleddin, Zohreh Rahnema, Mohammad Mohammadi, Elham Izady, Mehrdad Meshkinfam, Ergin Öpengin & Fatemeh Nemati. 2020. Turkic languages in the Atlas of the Languages of Iran (ALI). *Turkic Languages 24*(2). 290–308.
Brown, Cecil H., Eric W. Holman, Søren Wichmann & Viveka Velupillai. 2007. Automated classification of the world's languages: A description of the method and preliminary results. *STUF – Language Typology and Universals 61*. 285–308.
Bulut, Christiane. 2014. Turkic varieties in West Iran and Iraq: Representatives of a South Oghuz dialect group? In Heidi Stein (ed.), *Turkic in Iran: Past and present*, 15–99. Wiesbaden: Harrassowitz.
Bulut, Christiane. 2016. Convergence and variation in the Turkic varieties of Iran: Examples from Qashqâ'î. In Éva Á Csató, Lars Johanson, András Róna-Tas & Bo Utas (eds.), *Turks and Iranians: Interactions in language and history* (Turcologica 105), 235–82. Wiesbaden: Harrassowitz.
Burrough, Peter A., Rachael A. McDonnell & Christopher D. Lloyd. 2015. *Principles of Geographical Information Systems*. Oxford: Oxford University Press.
Cathcart, Chundra A. 2015. *Iranian dialectology and dialectometry*. Berkeley: University of California, Berkeley dissertation.
Chan, Eugene. 1998–2018. *Numeral systems of the world's languages*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://mpi-lingweb.shh.mpg.de/numeral (14 January 2022).
Christensen, Arthur. 1930/1935. *Contributions à la dialectologie iranienne*. Copenhagen: Levin & Munksgaard.
Durkin-Meisterernst, Desmond. 2004. *Dictionary of Manichaean Middle Persian and Parthian*. Turnhout: Brepols.
Dyen, Isidore. 1962. The lexicostatistically determined relationship of a language group. *International Journal of American Linguistics 28*(3). 153–61.
Erfani, Parisa. 2012. *Azeri Morphosyntax: The influence of Persian on a Turkic language*. Vancouver: Simon Fraser University thesis.
Fattāh, Ismaïl Kamandār. 2000. *Les dialectes du kurde méridional*. Louvain: Peeters.
Fattahi, Mehdi, et al. 2018. Language distribution in Kermanshah Province, Iran. In Erik Anonby, Mortaza Taheri-Ardali, et al., *Atlas of the Languages of Iran (ALI)*. http://iranatlas.net/module/language-distribution.kermanshah (14 January 2022).
[GCRC] Geomatics and Cartographic Research Centre. 2006–2021. *Nunaliit Atlas Framework*. Ottawa: GCRC, Carleton University. http://nunaliit.org (14 January 2022).
Gheitasi, Mojtaba, et al. 2017. Language distribution in Ilam Province, Iran. In Erik Anonby, Mortaza Taheri-Ardali, et al., *Atlas of the Languages of Iran (ALI)*. http://iranatlas.net/module/language-distribution.ilam (14 January 2022).
Haspelmath, Martin & Uri Tadmor (eds.). 2009. *Loanwords in the world's languages: A comparative handbook*. Berlin/New York: Mouton de Gruyter.
Hassandoust, Mohammad. 2011. *Farhang-e tatbiqi-mowzu'i-ye zabānhā va guyeshhā-ye irāni-ye now [Comparative-thematic dictionary of New Iranian languages and dialects]*. Tehran: Academy of Persian Language and Literature.
Hayes, Amos, Peter Pulsifer & J.-P. Fiset. 2014. The Nunaliit cybercartographic framework. In D. R. Fraser Taylor & Tracey P. Lauriault (eds.), *Developments in the theory and practice of cybercartography: Applications and Indigenous mapping*, 129–40. Amsterdam: Elsevier.
Hayes, Amos & D. R. Fraser Taylor. 2019. Developments in the Nunaliit cybercartographic data management platform. In D. R. Fraser Taylor, Erik Anonby & Kumiko Murasugi (eds.), *Further developments in the theory and practice of cybercartography: International dimensions and language mapping*, 205–18. Amsterdam: Elsevier.
Hourcade, Bertrand, Hubert Mazurek, Mohammad-Hosseyn Papoli-Yazdi, Mahmoud Taleghani, et al. 2012. *Irancarto*. Paris and Tehran: CNRS/University of Tehran. http://www.irancarto.cnrs.fr (14 January 2022).
[ISC] Iranian Statistics Centre. 2011/2016. Sālnāmeh-ye Āmār-e Irān [Iran Statistical Annals]. Tehran: Ministry of the Interior, Iranian Statistics Centre. http://www.amar.org.ir (14 January 2019).
Izady, Michael Mehrdad. 2013. *Linguistic composition of Iran (map)*. http://gulf2000.columbia.edu/images/maps/Iran_Languages_lg.png (14 January 2022).
Kıral, Filiz. 2000. Reflections on -mis in Kalaj. In Lars Johanson & Bo Utas (eds.), *Evidentials: Turkic, Iranian and neighbouring languages*, 89–102. Berlin/New York: Mouton de Gruyter.
Leitner, Bettina, Erik Anonby, Dina El Zarka & Mortaza Taheri-Ardali. 2021. A first description of Arabic on the south coast of Iran: The Arabic dialect of Bandar Moqām, Hormozgan. *Journal of Semitic Studies 66*(1). 215–61. https://doi.org/10.1093/jss/fgaa040.
MacKenzie, David N. 1971. *A concise Pahlavi dictionary*. Oxford: Oxford University Press.
Mann, Oskar. 1910. *Die Mundarten der Lur Stämme im südwestlichen Persien*. Berlin: Georg Reimer.
Mohammadirad, Masoud, Erik Anonby, Jaffer Sheyholislami, Mortaza Taheri-Ardali, Amos Hayes, J.-P. Fiset, Robert Oikle, Adam Stone, Sheema Rezaei, Emily Wang, Pegah Nikravan, Parisa Sabethemmatabadi, Ali Ghaharbeighi, Partow Mohammadi, Ayat Tadjalli, Nima Kiani, Laura Salisbury, et al. 2016. Language distribution in Kordestan Province, Iran. In Erik Anonby, Mortaza Taheri-Ardali, et al., *Atlas of the Languages of Iran (ALI)*. http://iranatlas.net/module/language-distribution.kordestan (14 January 2022).
Mohebbi Bahmani, Hassan, Ali Rashidi, Erik Anonby, et al. 2015. Language distribution in Hormozgan Province, Iran. In Erik Anonby, Mortaza Taheri-Ardali, et al., *Atlas of the Languages of Iran* (ALI). http://iranatlas.net/module/language-distribution.hormozgan (14 January 2022).

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

[NCC] National Cartographic Centre. 2015. *Pāygāh-e melli-ye nāmhā-ye ǰoghrāfiāi-ye Irān [National database of geographic names of Iran]*. Tehran: Ministry of the Interior, National Cartographic Centre. http://gndb.ncc.org.ir (14 January 2019).

Nemati, Fatemeh, Shakiba Ghasemi, Erik Anonby, et al. 2017. Language distribution in Bushehr Province, Iran. In Erik Anonby, Mortaza Taheri-Ardali, et al., *Atlas of the Languages of Iran (ALI)*. http://iranatlas.net/module/language-distribution.bushehr (14 January 2022).

Okati, Farideh, Erik Anonby, Abbas Ali Ahangar & Carina Jahani. 2010. Natural processes in Sistani Persian of Iran. *Iranian Journal of Applied Language Studies 2*(1). 93–120.

Oranskij, Iosif M. 1979. *Les langues iraniennes*. Paris: Klincksieck.

Orywal, Erwin (ed.). 1988. Karte 10, Vorderer Orient: Sprachen und Dialekte [Map 10, Middle East: Languages and Dialects]. Tübinger Atlas des Vorderen Orients (TAVO), series A, vol. 8. Wiesbaden: Reichert.

Persian Academy (Farhangestān-e Zabān-hā va Guyesh-hā-ye Irāni). 2009. *Rāhnemā-ye gardāvari-ye guyesh-hā [Guide to the collection of dialects]*. Tehran: Persian Academy. http://www.persianacademy.ir/UserFiles/File/Gouyesh.pdf (1 July 2017).

Peterson, Joseph H. 1995. *Dictionary of most common Avesta words*. http://www.avesta.org/avdict/avdict.htm (14 January 1995).

Roostanet 2016 = Rural Network of Iran. 2016. *Roostanet*. Tehran: Rural Network of Iran. http://map.roostanet.com (14 January 2019).

Sāsānpur, Farzāneh. 1393/2014. *Chahar Mahal va Bakhtiari. In Dāneshnāmeh-ye ǰahān-e eslām [Encyclopedia of the World of Islam] 12*. Tehran: Islamic Encyclopedia Foundation. http://rch.ac.ir/article/Details/9252 (4 July 2019).

Schmitt, Rüdiger (ed.). 1989. *Compendium linguarum iranicarum*. Wiesbaden: Reichart.

Schreiber, Laurentia, Geoffrey Haig, Mortaza Taheri-Ardali & Erik Anonby. 2021. Contact-induced change in the morphosyntax of Turkic in Boldaji, Chahar Mahal va Bakhtiari Province, Iran. *Turkic Languages 25*(2). 210–242.

Schreiber, Laurentia, Mortaza Taheri-Ardali, Erik Anonby & Geoffrey Haig. 2017. Contact-induced change in Irano-Turkic morphosyntax. Conference on Language Contact and Language Change in Western Asia, Goethe University Frankfurt, March 10–12, 2017.

Schwartz, Martin. 1982. 'Blood' in Sogdian and Old Iranian. *Monumentum Georg Morgenstierne II*. Leiden: Brill. 189–96.

Stilo, Don. 2016. The Araxes-Iran linguistic area with a concentration on Kurdophone varieties. Paper presented at the 3rd International Conference on Kurdish Linguistics (ICKL3), University of Amsterdam, August 25–26, 2016.

Swadesh, Morris. 1950. Salish internal relationships. *International Journal of American Linguistics 16*(4). 157–67.

Swadesh, Morris. 1971. *The origin and diversification of language*, ed. Joel Sherzer. Chicago: Aldine.

Taheri-Ardali, Mortaza. 2015. Guyesh-e Bakhtiāri: zavāl yā baghā? Negāhi digar be kārbord-e guyesh-e Bakhtiāri, guneh-ye Ardal [Bakhtiari dialect: Decline or survival? A re-examination of the usage of the Bakhtiari dialect, Ardal variety]. Paper presented at the International Conference on Linguistic Heritage, Tehran, October 20–21, 2015.

Taheri-Ardali, Mortaza. 2017a. Salient linguistic features and isoglosses in Chahar Mahal va Bakhtiari Province, Iran. 1st North American Conference in Iranian Linguistics (NACIL1), Stony Brook University, New York, April 28–30, 2017.

Taheri-Ardali, Mortaza. 2017b. A first account of the languages in western Hormozgan. 7th International Conference on Iranian Linguistics (ICIL7), Lomonosov Moscow State University and the Institute of Linguistics of the Russian Academy of Sciences, Moscow, August 28–30, 2017.

Taheri-Ardali, Mortaza. 2020. Ostān-e chāhār mahāl va bakhtiāri dar atlas-e zabānhā-ye irān: Raveshshenāsi-ye pazhuhesh va parākandegi-ye zabāni [Chahar Mahal va Bakhtiari Province in the Atlas of the Languages of Iran: Research methodology and language distribution]. *Comparative Linguistic Research (Iran) 10*. 47–69. https://doi.org/10.22084/rjhll.2020.20969.2004

Taheri-Ardali, Mortaza, Erik Anonby, Adam Stone, Amos Hayes, J.-P. Fiset, Robert Oikle, Laura Salisbury, Mahnaz Talebi-Dastenaee, Maryam Amani-Babadi, et al. 2015. Language distribution in Chahar Mahal va Bakhtiari Province, Iran. In Erik Anonby, Mortaza Taheri-Ardali et al., *Atlas of the Languages of Iran (ALI)*. http://iranatlas.net/module/language-distribution.chahar_mahal_va_bakhtiari (14 January 2022).

Taheri-Ardali, Mortaza & Erik Anonby. 2019. Some salient features of the Bakhtiari language in Chahar Mahal va Bakhtiari Province. *Proceedings of the National Conference on Bakhtiari Language, Culture and Literature, Payam-e Noor University of Shahr-e Kord, 9 May, 2018*, 993–1001. Tehran: Payam-e Noor University. http://conference.pnu.ac.ir/ChaharMahal-bakhtiari/post.aspx?id=1312 (14 January 2020).

Taheri-Ardali, Mortaza, Erik Anonby, Mahnaz Talebi-Dastenaee, Peyman Pishyar, Fatemeh Shahverdi, Elham Hasanpour, Reza Rezvani-Borujeni, et al. 2020. Chahar Mahal va Bakhtiari linguistic data. In Erik Anonby, Mortaza Taheri-Ardali, et al. (eds.). *ALI Dataverse*. Toronto: Scholars Portal Dataverse. https://doi.org/10.5683/SP2/FVLDLZ.

Taheri-Ardali, Mortaza, Erik Anonby, Amos Hayes, et al. 2021. Atlas-e barkhatt-e zabanhā-ye Irān: tarāhi, raveshshenāsi va natāyej-e avvalieh [The online Atlas of the Languages of Iran: Design, methodology and initial results.] *Language Related Research 12*(2). 231–291.

Talebi-Dastenaee, Mahnaz & Erik Anonby. Forthcoming. Raji (Barzoki) lexicon. In Paul Heggarty & Cormac Anderson (eds.), *Comparison of basic lexicon*. Jena: Max Planck Institute for the Science of Human History. http://cobl.info (14 January 2022).

Talebi-Dastenaee, Mahnaz, Hamideh Poshtvan & Erik Anonby. Under consideration. Raji in contact with Persian: Contrasting paths to convergence in two dialects.

Taylor, D. R. Fraser. 2003. The concept of cybercartography. In M. P. Peterson (ed.), *Maps and the Internet*, 405–520. Amsterdam: Elsevier.

Taylor, D. R. Fraser. 1997. Maps and mapping in the information era. Keynote address to the 18th International Cartographic Association Cartographic Conference, June 23, 1997, Stockholm, Sweden. In L. Ottoson (ed.), Proceedings of the International Cartographic Conference 97(1), 1–10. Gavle: Swedish Cartographic Society.

Taylor, D.R. Fraser. 2005. The theory and practice of cybercartography: An introduction. In D. R. Fraser Taylor (ed.), *Cybercartography: Theory and practice*, 1–14. Amsterdam: Elsevier.

Taylor, D. R. Fraser & Tracey P. Lauriault (eds.). 2014. *Developments in the theory and practice of cybercartography: Applications and indigenous mapping*. Amsterdam: Elsevier.

Taylor, D. R. Fraser, Erik Anonby & Kumiko Murasugi (eds.). 2019. *Further developments in the theory and practice of cybercartography: International dimensions and language mapping*. Amsterdam: Elsevier.

Windfuhr, Gernot. 1988. Baḵtīārī tribe: The Baḵtīārī dialect. *Encyclopædia Iranica* 3.5, 553–60. New York: Center for Iranian Studies, Columbia University. http://www.iranicaonline.org/articles/baktiari-tribe#pt2 (14 January 2019).

Windfuhr, Gernot. 2009. Dialectology and topics. In Gernot Windfuhr (ed.), *The Iranian languages*, 5–42. London/New York: Routledge.

Zhukovskij, Valentin A. (ed. S. F. Ol'denburg). 1923. *Materialy dlya izucheniya persidskikh narechij, pt. 3: Narechie bakhtiyarov chekhar-leng i kheftleng [Materials for the study of Persian dialects, part 3: The dialect of the Čahārlang and Haftlang Bakhtiari]*. Petrograd.

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

Appendix 1.
Lexical data set (archived data set with full accompanying metadata is available at: https://doi.org/10.5683/SP2/FVLDLZ)

language variety/site	language grouping	variety code	1. 'hair (of head)'	2. 'eye'	3. 'nose'	4. 'ear'	5. 'mouth'	6. 'tongue'	7. 'throat (inside)'	8. 'arm'	9. 'hand'
Persian, Standard (reference variety)	Persian	PRS	*mu*	*češm*	*bini, damāġ*	*guš*	*dahān*	*zabān*	*galu*	*dast*	*dast*
Masjed Soleymān (reference variety)	Bakhtiari	MJS	*mi*	*tey*	*noft*	*guš*	*dohoʊn*	*zoʊn*	*geli*	*čel*	*dast*
Sar Āqā Seyyed	Bakhtiari	SAS	*mü*	*ti*	*noft*	*guš*	*dohaʊ*	*zaʊ*	*gelü, tešni*	*čel*	*pange*
Deh Now Soflā	Bakhtiari	DNO	*mi*	*tiye*	*noft*	*guš*	*dohaʊn, dahaʊn*	*zaʊn*	*gili*	*bāyi*	*panga*
Sepidāneh	Bakhtiari	SEP	*mi*	*tē*	*noft*	*guš*	*dohon*	*zōn*	*gili*	*dast*	*dast*
Loshtar Gorui	Bakhtiari	LGI	*mi*	*tiye*	*neft*	*guš*	*diyaʊn*	*zaʊn*	*gili*	*das*	*das*
Fārsān	Bakhtiari	FRS	*mi*	*tiye*	*noft*	*guš*	*dohun*	*zeʊun*	*gili*	*dast*	*dast*
Juneqān (Bakhtiari)	Bakhtiari	JNB	*mi*	*ti, tiya*	*noft*	*guš*	*dohun*	*zun*	*gili*	*dast*	*dast*
Ardal	Bakhtiari	ARD	*mi*	*tiye*	*neft*	*guš*	*dun*	*zun*	*gili*	*dast*	*dast*
Shalamzār	Bakhtiari	SHL	*mi*	*tiye*	*neft*	*guš*	*dohun*	*zeʊun*	*gili*	*dast*	*dast*
Boldāji (Bakhtiari)	Bakhtiari	BLB	*mi*	*tiya*	*noft*	*guš*	*duhun*	*zobun*	*gili*	*dast*	*dast*
Lordegān	Bakhtiari	LDG	*mel*	*tiya*	*noft*	*guš*	*dun, dahan, čil*	*zoʊn*	*gili*	*das*	*das*
Chilteh Duderā	Bakhtiari	CHT	*mel*	*tiya*	*noft, fir*	*guš*	*čil*	*zeʊun*	*bot, xor, gili*	*dahs*	*dahs*
Cham Chang	Charmahali	CHC	*mu, mi*	*češm*	*damāġ, domāġ*	*guš*	*dahan*	*zebun*	*gelu, gili*	*dast*	*dast*
Shurāb Kabir (Charmahali)	Charmahali	SKC	*mu, mi*	*češ*	*damāġ, domāġ*	*guš*	*dahan*	*zebun*	*gilu, gelu*	*das*	*das*
Sheykh Shabān	Charmahali	SHS	*mi*	*tiye, tiya*	*puz, damāġ*	*guš*	*duhun*	*zubun*	*gili*	*dast*	*dast*
Arjenak	Charmahali	ARJ	*mi*	*češ*	*damāġ, bini*	*guš*	*dahan*	*zobun*	*gili*	*das*	*das*
Fath Ābād	Charmahali	FTA	*mi*	*češ*	*domāġ*	*guš*	*dahan*	*zebun*	*gili*	*dast*	*dast*
Hāruni	Charmahali	HAR	*mi*	*češ*	*damāġ, noft*	*guš*	*dahan*	*zebun*	*gili*	*dast*	*dast*
Shahr-e Kord	Charmahali	SKO	*mu*	*češ*	*domāġ, bini*	*guš*	*dahan*	*zobun*	*gilu*	*das*	*das*
Farrokh Shahr	Charmahali	FSH	*mü*	*češ*	*dumāġ*	*guš*	*dan*	*zebun*	*gülü*	*das*	*das*
Hafshejān	Charmahali	HAF	*mu, mi*	*češ*	*domāġ*	*guš*	*dahan*	*zebun, zabun*	*gilu, gili*	*das*	*das*
Borujen	Charmahali	BOR	*mu, gis*	*češ*	*domāġ*	*guš*	*dan*	*zebun*	*gilu*	*das*	*das*
Naqneh (Charmahali)	Charmahali	NQC	*gis*	*češ*	*domāġ*	*guš*	*dahn*	*zobun*	*gili, gilu*	*das*	*das*
Ben	Turkic	BEN	*tük*	*göz*	*burn*	*ġulāx*	*āġz*	*dil*	*boġāz*	*al*	*al*
Margh Malek	Turkic	MAM	*tik*	*gez*	*bur*	*ġulāġ*	*āġz*	*dil*	*boġāz*	*al*	*al*
Shurāb Kabir (Turkic)	Turkic	SKT	*ġel*	*göz*	*burni*	*ġulāx*	*āġz*	*dil*	*boġāz*	*al*	*al*
Shahr-e Kiān	Turkic	KIN	*tik*	*göz*	*burn*	*ġulāġ*	*āġz*	*dil*	*boġāz*	*al*	*al*
Juneqān (Turkic)	Turkic	JNT	*balig*	*göz*	*bun*	*ġulāġ*	*āġz*	*dil*	*bāġāz*	*al*	*al*
Boldāji (Turkic)	Turkic	BLT	*tik*	*gez*	*burn*	*ġulāġ*	*āġz*	*dil*	*boġāz*	*al*	*al*
Naqneh (Turkic)	Turkic	NQT	*tik, ġil*	*gez*	*burn*	*ġulāġ*	*āġz*	*dil*	*boġāz, xerrex*	*al*	*alayāx*
Sulegān	Turkic	SUL	*tik*	*gez*	*burn*	*ġulāġ*	*āġz*	*dil*	*xerrex*	*al*	*al*

variety	10. ‘finger’	11. ‘stomach (belly)’	12. ‘stomach (organ, human)’	13. ‘leg’	14. ‘knee’	15. ‘foot’	16. ‘bone’	17. ‘blood’	18. ‘urine’	19. ‘heart (organ)’
PRS	*angošt*	*šekam*	*me’de*	*pā*	*zānu*	*pā*	*ostoxān*	*xun*	*edrār, šāš*	*ġalb*
MJS	*angost*	*eškam*	*gade*	*leng*	*zuni*	*pā*	*hast, ostoxoun*	*hin*	*meste*	*del, qalb*
SAS	*kelek*	*ešgam*	*mehre*	*leng, tekerau*	*zaui*	*pā*	*ostoxau*	*hün*	*meste*	*qalu*
DNO	*angost, kelek*	*kom*	*del, mahde*	*pā, leng*	*zuui*	*pā*	*hast, ostoxaun*	*hin, xin*	*mesta*	*ġalb*
SEP	*angost*	*kom*	*mehde*	*pā*	*zuni*	*pā*	*ostoxun*	*hin*	*meste*	*ġalb*
LGI	*noxaun*	*kom*	*mehde*	*pā*	*zuni*	*pā*	*xas, ostoxaun*	*xi*	*mehse*	*ġalb, del*
FRS	*angošt*	*eškam*	*mahde*	*pā*	*zuni*	*pā*	*ostoxun*	*xin*	*meste*	*del*
JNB	*angost, angošt*	*tel, eškam, kom*	*mehda, mehde, komb*	*pā*	*zuni*	*pā*	*ostoxun*	*xin*	*šāš*	*ġalb*
ARD	*angost*	*kom*	*del, mehde*	*pā, leng*	*zuni*	*pā*	*estexun*	*xin*	*meste*	*ġalb*
SHL	*angošt*	*kom, del*	*mehde, mahde*	*pā*	*zuni*	*pā*	*ostoxun*	*xin*	*šāš*	*ġalb*
BLB	*angošt*	*kom*	*mahda*	*pā*	*zuni, zānu*	*pā*	*ostoxun*	*xin*	*šāš*	*del*
LDG	*kilič*	*kom*	*mehde, del*	*pā, tekerun*	*zuni*	*pā*	*ostoxun*	*xin*	*mehse*	*ġalb*
CHT	*čelik, kilič*	*kom*	*kom*	*pā*	*zuni*	*pā*	*hahs*	*xin*	*mehsa*	*qalb*
CHC	*angušt*	*šikam, del*	*mehde*	*pā*	*zāni*	*pā*	*ostoxun*	*xun, xin*	*šāš*	*ġalb*
SKC	*angošt, anguli*	*šikam*	*mehde*	*pā*	*zānu*	*pā*	*ostoxun*	*xun, xin*	*šāš*	*ġalb*
SHS	*panǰe*	*del*	*mehde*	*pā*	*zuni*	*pā*	*estexun*	*xin*	*meste*	*ġalb*
ARJ	*angušt, angošt*	*šikam*	*mede, made*	*pā*	*zāni*	*pā*	*ostoxun*	*xin*	*šāš, miz, meste*	*ġalb*
FTA	*angošt*	*del*	*mehde*	*pā*	*zuni*	*pā*	*ostoxun*	*xin*	*šāš*	*ġalb*
HAR	*angošt, anguli*	*šikam*	*mehde*	*pā*	*zuni, zānu*	*pā*	*ostoxun*	*xin, xun*	*šāš, edrār*	*ġalb*
SKO	*angoš, anguli*	*šikam, del*	*mehde*	*pā, leng*	*zānu*	*pā*	*ostoxun, ostoxum*	*xin*	*pišāb, šāš*	*ġalb*
FSH	*anguli*	*šikam, del*	*mahde*	*pā*	*zāni*	*pā*	*ostoxun*	*xün*	*pišāb, šāš*	*ġalb*
HAF	*angoš*	*šikam*	*mehde, mahde*	*pā*	*zānu, zuni*	*pā*	*estexun*	*xin*	*šāš*	*ġalb, ġabl*
BOR	*angošt*	*šikam*	*ma’de*	*pā, leng*	*zānu, zinu*	*pā*	*ostoxun*	*xin*	*šāš*	*ġalb*
NQC	*angošt, anguli*	*šikam*	*del*	*pā*	*zāni*	*pā*	*ostoxun*	*xin, xun*	*šāš*	*ġalb*
BEN	*birmāx*	*ġārn*	*ürag*	*ġič*	*diz*	*ġič*	*sümük*	*ġān*	*išamak*	*ġalb*
MAM	*burmāġ*	*ġārn*	*mehda*	*ġeč*	*diz*	*ġeč*	*simig*	*ġān*	*iššamāġ*	*irag, ġalb*
SKT	*birmāx*	*irag*	*mehda*	*ġeč*	*diz*	*ġeč*	*simik*	*ġān*	*pišāb*	*ġalb*
KIN	*burmāġ*	*ġārn, irag*	*mehde, irag*	*ġeč*	*diz*	*aġ*	*sömik*	*ġān*	*išamāġ*	*ġalb*
JNT	*burmāġ*	*irey*	*mehda*	*aġ, ayaġ*	*diz*	*aġ, ayaġ*	*simig*	*ġān*	*išamāġ*	*irey, ġalb*
BLT	*burmāġ, dernāġ*	*ġārn*	*me’de, mehde*	*ġeč*	*diz*	*ġeč*	*simik*	*ġān*	*išamāġ, pišāb*	*ġalb*
NQT	*burmāx*	*ġārn*	*mehda*	*ġič*	*diz*	*ġič*	*sinik, sünük*	*ġān*	*čorramag*	*irag*
SUL	*burmāġ*	*ġārun*	*mahda*	*ġič*	*diz*	*ġič*	*sinik*	*ġān*	*iššamag*	*ġalb*

variety	20. ‘liver’	21. ‘man’	22. ‘husband’	23. ‘woman’	24. ‘wife’	25. ‘child’	26. ‘boy’	27. ‘son’	28. ‘girl’	29. ‘daughter’
PRS	*ǰegar*	*mard*	*šaʋhar*	*zan*	*zan*	*bače, bačče*	*pesar*	*pesar*	*doxtar*	*doxtar*
MJS	*ǰiyar*	*piyā*	*meyre*	*zeyne*	*zeyne*	*bače*	*kor*	*kor*	*doʋdar, dohdar*	*doʋdar, dohdar*
SAS	*ǰiyar*	*piyā*	*mire*	*zine*	*zine*	*bače*	*kor*	*kor*	*dorar, doʋar*	*dorar, doʋar*
DNO	*ǰiyar*	*piyā*	*mira*	*zine, keyʋenu*	*zine*	*bače, aʋlād*	*kor, ǰeġela*	*kor*	*dodar*	*dodar*
SEP	*ǰeyar*	*mire*	*mire*	*zine*	*zine*	*bače*	*kor*	*kor*	*doʋar*	*doʋar*
LGI	*ǰiyar*	*meyre, piyā*	*meyre*	*zeyne*	*zeyne*	*bahče*	*kor*	*kor*	*doder*	*doder*
FRS	*ǰigar*	*mire*	*mire*	*zine*	*zine*	*bače*	*kor*	*kor*	*doʋar*	*doʋar*
JNB	*ǰiyar*	*mira*	*mira*	*zina*	*zina, ayāl*	*bače*	*ǰeġela*	*kor*	*doʋar*	*doʋar*
ARD	*ǰiyar*	*mire*	*mire*	*zine*	*zine*	*bače, aʋlād*	*kor*	*kor*	*doʋar*	*doʋar*
SHL	*ǰigar*	*mire, merd*	*mire, merd*	*zine*	*zine*	*bačče*	*kor, ǰeġele*	*kor*	*doxtar*	*doxtar*
BLB	*ǰigar*	*merd*	*šuʋar, mire*	*zan*	*zan*	*bača*	*kor, ǰeġele*	*kor*	*doxtar, doʋar*	*doxtar, doʋar*
LDG	*ǰigar*	*mire, piyā*	*mire*	*zine*	*zine*	*bače*	*kor*	*kor*	*doʋar*	*doʋar*
CHT	*geǰar*	*piyā*	*mira*	*zina*	*zina*	*bača*	*kor*	*kor*	*duʋar*	*duʋar*
CHC	*ǰigar*	*mard*	*šiʋar*	*zan*	*zan*	*beče*	*pesar*	*pesar*	*doxtar*	*doxtar*
SKC	*ǰigar*	*mard*	*šuʋar*	*zan*	*zan*	*beče*	*pesar, ǰeġele*	*pesar, ǰeġele*	*doxtar*	*doxtar*
SHS	*ǰigar*	*merd*	*mire*	*zan*	*zan*	*bačče*	*kuʋak, ǰeġele, kor*	*kuʋak*	*dodar*	*dodar*
ARJ	*ǰigar*	*mard, piyā*	*šiʋar*	*zan*	*hamsar*	*beče*	*pesar, ǰeġele*	*pesar, beče*	*doxtar*	*doxtar*
FTA	*ǰigar*	*mard*	*šuʋar*	*zan*	*zan*	*bače*	*ǰeġele*	*pesar*	*doxtar*	*doxtar*
HAR	*ǰigar*	*mard*	*šuʋar*	*zan*	*zan*	*beče*	*pesar, ǰeġele*	*pesar*	*doxtar*	*doxtar*
SKO	*ǰigar*	*mard*	*šuʋar*	*zan*	*zan*	*beče*	*pesar, piyā*	*pesar*	*doxtar*	*doxtar*
FSH	*ǰigar*	*mard*	*šuʋar*	*zan*	*zan, zayfe*	*beče*	*ǰeġele*	*ǰeġele*	*doxtar, māyne*	*doxtar*
HAF	*ǰigar*	*mard*	*šuʋar*	*zan*	*zan*	*bačče*	*pesar*	*pesar*	*doxtar*	*doxtar*
BOR	*ǰigar*	*mard*	*šuʋar*	*zan*	*zan, zayfe*	*beče*	*kor, pesar, ǰeġele*	*kor, pesar*	*doxdar*	*doxdar*
NQC	*ǰigar*	*mard*	*šuʋar*	*zayife*	*zayife, ayāl, zan*	*beče, ǰeġele*	*pesar, ǰeġele*	*pesar*	*doxdar, mādine*	*doxdar, mādine*
BEN	*ǰigar*	*kiša*	*ar*	*ārʋād*	*ārʋād*	*ušāx*	*oġlān*	*oġlān*	*ġiz*	*ġiz*
MAM	*ǰigar*	*kiši*	*ar*	*arʋād, ārʋād*	*arʋāye, ārʋāye, arʋāde*	*ušāġ, išāġ*	*oġlān*	*oġlān*	*ġez*	*ġez*
SKT	*ǰigar*	*kiše*	*kiše*	*ārʋād*	*hamsar*	*ušāġ*	*oġlān*	*oġlān*	*ġez*	*ġez*
KIN	*ǰigar*	*kiši*	*ar*	*ārʋād*	*ārʋād*	*ušāġ*	*oġlān, oġol*	*oġlān, oġol*	*ġez*	*ġez*
JNT	*ǰiyar*	*ar*	*ar*	*oyra*	*oyra*	*ušāġ*	*oġel*	*oġel*	*ġez*	*ġez*
BLT	*ǰigar*	*kiši*	*arr*	*ārʋād*	*ārʋād*	*ušāġ*	*ǰeġela*	*oġlān*	*ġez*	*ġez*
NQT	*ǰigar*	*kiše*	*ar*	*zeyfa*	*ārʋād*	*uššāx*	*oġol*	*oġol*	*ġiz*	*ġiz*
SUL	*ǰigar*	*kiše*	*ar*	*zeyfa*	*ārʋād, zeyfa*	*uššāġ*	*oġol*	*oġol*	*ġiz*	*ġiz*

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

variety	30. 'groom'	31. 'bride'	32. 'dog'	33. 'cat'	34. 'wolf'	35. 'fox'	36. 'fish'	37. 'scorpion'	38. 'tree'	39. 'wood (substance)'	40. 'branch'
PRS	*dāmād*	*arus*	*sag*	*gorbe*	*gorg*	*rubāh*	*māhi*	*aġrab*	*deraxt*	*čub*	*šāxe*
MJS	*dovā*	*arus, bahig*	*say*	*gorbe*	*gorg*	*ruvā*	*māhi*	*gaždin*	*dār*	*ču*	*lešk*
SAS	*dovā*	*behüg*	*say*	*gorve*	*gorg*	*ruvā*	*māhi*	*gaždün*	*deraxt*	*ču*	*lešk, lahk*
DNO	*dovā*	*behig*	*say*	*gorbe*	*gorg*	*ruvā*	*moi*	*gaždin*	*deraxt*	*ču*	*lešk, lak*
SEP	*dovā*	*arus*	*sag*	*gorbe*	*gorg*	*ruvā*	*mohi*	*gaždin*	*deraxt*	*ču*	*lešk*
LGI	*dumā*	*arus, behig*	*sag*	*gulu, gulubiš*	*gorg*	*ruvā*	*moi*	*gādim*	*derahd*	*ču*	*lešk*
FRS	*dumā*	*āris*	*sag*	*gorbe*	*gorg*	*rubā*	*māhi*	*každin*	*daraxt*	*ču*	*šāxe, lešk*
JNB	*dumā*	*arus, behig*	*sag*	*gorba, gorbe*	*gorg*	*ruvā*	*māhi*	*aġrab, každom*	*deraxt, daraxt*	*ču*	*šāxa, šāxe*
ARD	*dumā*	*arus, beheyg*	*sag*	*gorbe*	*gorg*	*ruvah*	*māi*	*gādim*	*deraxt*	*ču*	*lešk, šāxe*
SHL	*dumā*	*āris*	*sag*	*gorbe*	*gorg*	*ruvā*	*māyi*	*gaždin*	*daraxt*	*ču*	*šāxe, leške*
BLB	*dumā*	*arus*	*sag*	*gorba*	*gorg*	*rubā*	*māhi*	*aġrab*	*daraxt*	*čub, ču*	*šāxa*
LDG	*dumā*	*arus, beyig*	*sag, kotu*	*gulu*	*gorg*	*ruvah*	*moyi*	*gādim*	*darax*	*ču*	*lešxa*
CHT	*dumā*	*bavig*	*ketu*	*gelu*	*gorg*	*ruvah*	*movi*	*dindarakul, gādim*	*dār*	*ču*	*lešxa, laxa*
CHC	*dumād, dāmād*	*arus*	*sag*	*gorbe*	*gorg*	*rubā*	*māhi*	*aġrab*	*deraxt*	*čub, čuġ*	*šāxxe*
SKC	*dumād, dāmād*	*arus*	*sag*	*gorbe*	*gorg*	*rubā*	*māhi*	*aġrab*	*deraxt*	*čuġ*	*šāxe*
SHS	*dumā*	*āris*	*sag*	*gorbe*	*gorg*	*rubā*	*māhi*	*aġrab*	*deraxt*	*ču*	*šāxe*
ARJ	*dumād*	*arus*	*sag*	*gorbe*	*gorg*	*rubā, ruvā*	*māhi*	*každom, gižd̈om, aġrab*	*dār, darax*	*čuġ, čub*	*šāxe, par*
FTA	*dumād*	*arus*	*sag*	*gorbe*	*gorg*	*rubā*	*māhi*	*aġrab*	*daraxt*	*čuġ*	*šāxe*
HAR	*dumād*	*arus*	*sag*	*gorbe*	*gorg*	*rubā*	*māhi*	*aġrab*	*daraxt*	*čuġ*	*šāxe*
SKO	*dumād*	*arus*	*sag*	*gorbe, pišuli*	*gorg*	*rubā*	*māy, māyi*	*aġrab*	*darax*	*čuġ, čub*	*šāxe*
FSH	*dumād*	*arus*	*sag*	*gorbe*	*gorg*	*rubā*	*māy*	*aġrab*	*derax*	*čuġ*	*šāxe*
HAF	*dumād*	*arus*	*sag*	*gorbe*	*gorg*	*rubā*	*māhi*	*aġrab*	*daraxt*	*čuġ*	*šāxxe*
BOR	*dumād*	*arus*	*sag*	*gorbe*	*gorg*	*rubā*	*māy*	*aġrab, každom*	*deraxt*	*čuġ*	*šāxe*
NQC	*dumād*	*arus*	*sag*	*gorbe*	*gorg*	*rubā*	*māy*	*aġrab*	*derax*	*čuġ*	*šāxe*
BEN	*kürakan*	*galen*	*köpag, it*	*püšüg*	*ġurt*	*tilka*	*bālöx*	*axrab*	*āġāǰ*	*āġāǰ*	*šāġġā*
MAM	*kirakan*	*galen*	*kepag*	*pišig*	*ġurt*	*tilki, ruvā*	*moyi*	*geždin, aġrab*	*deraxt*	*āġāǰ*	*šāġġa, šāxa*
SKT	*dāmād*	*arus*	*köpag*	*pišig*	*ġurt*	*tülke*	*bālux*	*aġrab*	*daraxt*	*āġāǰ*	*šāxa*
KIN	*dumād*	*arus*	*it*	*pišig*	*ġurt*	*tilki, rubā*	*māyi*	*aġrab*	*deraxt*	*āġāǰ*	*tarka, šāxa*
JNT	*dumā*	*gahle*	*it*	*mali*	*ġurt*	*ruvā*	*māyi*	*gaždim, každim*	*daraxt*	*āġāǰ*	*šāġġa*
BLT	*dumād, dumā*	*galin*	*kepag*	*pišug*	*ġurt*	*rubā*	*māhi*	*dombekul, aġrab*	*deraxt*	*āġāǰ*	*šāxa*
NQT	*kirakan*	*galen*	*it, köpag*	*gorbe*	*ġurt*	*tilke*	*bālox*	*aġrab*	*āġāǰ*	*āġāǰ*	*šāxā*
SUL	*kirakan*	*galin*	*it, kepag*	*gurbā*	*ġurt*	*tilke*	*bāloġ*	*aġrab*	*āġāǰ*	*āġāǰ*	*šāxā*

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

variety	41. ‘stick’	42. ‘leaf’	43. ‘sun’	44. ‘day’	45. ‘night’	46. ‘star’	47. ‘water’	48. ‘winter’	49. ‘house’	50. ‘rice’
PRS	*čub, čubdasti*	*barg*	*xoršid*	*ruz*	*šab*	*setāre*	*āb*	*zemestān*	*xāne*	*berenǰ*
MJS	*kalāk*	*par*	*aftɑʊ*	*ruz*	*šɑʊ*	*āstāre*	*ɑʊ*	*zemestoʊn*	*hoʊne*	*berenǰ*
SAS	*ču, gorz*	*pahr*	*aftɑy*	*ruz*	*šɑy*	*āstāre*	*ɑy*	*zemestɑʊ*	*hɑʊe*	*berenǰ*
DNO	*ču, tarka, gorz*	*pahr*	*aftɑʊ*	*ruz*	*šɑʊ*	*āstāre*	*ɑʊ*	*zemestɑʊn*	*hoʊɑ*	*berenǰ*
SEP	*gorz, tarke, kalāk*	*pahr*	*aftɑʊ*	*ruz*	*šɑʊ*	*setāre, āstāre*	*ɑʊ*	*zemeston*	*hune*	*berenǰ*
LGI	*ču, gorz*	*pahr*	*aftɑʊ*	*ruz*	*šɑʊ*	*āsāre*	*ɑʊ*	*zemesɑʊn*	*hɑʊne*	*berenǰ*
FRS	*ču, tarke, gorz*	*pahr*	*āftɑʊ*	*ruz*	*šɑʊ*	*āstāre, setāre*	*ɑʊ*	*zemestun*	*huna*	*berenǰ*
JNB	*ču*	*pahr*	*aftɑʊ*	*ruz*	*šɑʊ*	*setāre, āstāre*	*ɑʊ*	*zemestun*	*hoʊne*	*berenǰ*
ARD	*gorz, ču*	*pahr*	*aftɑʊ*	*ruz*	*šɑʊ*	*setāre, āstāre*	*ɑʊ*	*zemestun*	*hune*	*berenǰ*
SHL	*ču dasti, gorz*	*pahr*	*oftɑʊ*	*ruz*	*šɑʊ*	*setāre, āstāre*	*ɑʊ*	*zemestun*	*xune*	*berenǰ*
BLB	*čub, ču*	*par*	*oftɑʊ*	*ruz*	*šɑʊ*	*setāre*	*ɑʊ*	*zemestun*	*xuna*	*berenǰ*
LDG	*čudas, gorz*	*par*	*aftoʊ*	*ruz*	*šoʊ*	*setāre, āsāre*	*oʊ*	*zemesun*	*huna*	*berenǰ*
CHT	*ġalāk, galāk*	*pahr*	*aftɑʊ*	*ruz*	*šɑʊ*	*ostāra*	*ɑʊ*	*zemehsun*	*huna*	*berenǰ*
CHC	*čub, čuġ*	*barg*	*oftɑʊ*	*ruz*	*šab, šey*	*setāre*	*ɑʊ*	*zemestun*	*xune*	*berenǰ*
SKC	*čuġ*	*barg, pahr*	*oftɑʊ*	*ruz*	*šab, šey*	*setāre*	*ɑʊ*	*zemastun*	*xune*	*berenǰ*
SHS	*ču, ču dasti*	*barg*	*oftɑʊ*	*ruz*	*še*	*āstāre, setāre*	*ɑʊ*	*zemestun*	*hune*	*berenǰ*
ARJ	*čomāġ, gorz, pāye*	*par*	*oftoʊ*	*ruz*	*šoʊ, šey*	*setāre*	*oʊ*	*zemastun*	*xune*	*berey*
FTA	*čuġ*	*pahr*	*oftɑʊ*	*ruz*	*šɑʊ*	*setāre*	*ɑʊ*	*zemestun*	*xune*	*berenǰ*
HAR	*čuġ, tarke*	*pahr*	*oftɑʊ*	*ruz*	*šɑʊ*	*setāre*	*ɑʊ*	*zemestun*	*xune*	*berenǰ*
SKO	*gorz, čuġ*	*barg*	*oftoʊ, āftāb*	*ruz*	*šum, šab*	*setāre*	*oʊ, āb*	*zemastun*	*xune*	*berey, berenǰ*
FSH	*čuġdasti*	*balg, balge*	*oftoʊ*	*ruz*	*šö*	*setāre*	*oʊ*	*zemastun*	*xune, öšā*	*berenǰ*
HAF	*asā*	*balg*	*ofto*	*ruz*	*še, šab*	*setāre*	*o*	*zemastun, zamastun*	*xune*	*berenǰ*
BOR	*čuġ, gorz*	*barg*	*oftoʊ*	*ruz*	*šɑʊ, šab*	*setāre*	*oʊ*	*zemastun, zemestun*	*xune*	*berenǰ*
NQC	*čuġ*	*balg, barg*	*oftɑʊ*	*ruz*	*šey*	*setāre*	*ɑʊ, āb*	*zemessun, zemestun*	*xune*	*bereyn, berenǰ*
BEN	*alāġāǰa*	*par*	*gün*	*günüsün*	*geǰa*	*ulduz*	*su*	*ġiš*	*öʊ*	*düga*
MAM	*āġāǰ, alāġāǰi*	*pahr, barg*	*gin*	*giniz*	*geǰa*	*setāra, olduz, āstāra*	*su*	*ġeš*	*eʊ*	*digi*
SKT	*alāġāǰe*	*barg*	*gün*	*ruz*	*šab*	*setāre*	*su*	*ġiš*	*eʊ*	*dügü*
KIN	*āġāǰ, tarka*	*barg, pahr*	*gün*	*gün, gündüz*	*geǰa*	*ulduz*	*su*	*ġeš*	*öy*	*digi*
JNT	*āġāǰ, alāġāǰi*	*barg*	*gün*	*gün*	*geyǰa, šām*	*olduz*	*su*	*ġeš*	*ey*	*berenǰ*
BLT	*āġāǰ, alāġāǰi*	*par, barg*	*gin*	*ginizin, gindiz*	*geǰa*	*ildiz*	*su*	*ġeš*	*ey*	*digi*
NQT	*āġāǰ, alāġāǰi*	*barg*	*gin, gün*	*ginnez, günnüz*	*geǰa*	*ulluz*	*su*	*ġiš*	*ey*	*digi*
SUL	*alāġāǰe*	*barga*	*gin*	*ginniz*	*geǰa*	*ulluz*	*su*	*ġiš*	*ɑʊ, obā*	*berenǰ*

variety	51. ‘egg (of chicken)’	52. ‘walnut’	53. ‘hungry’	54. ‘thirsty’	55. ‘bitter’	56. ‘fresh (e.g., milk)’	57. ‘long (thing)’	58. ‘dry’	59. ‘big’	60. ‘red’
PRS	*toxmemorġ*	*gerdu*	*gorosne*	*tešne*	*talx*	*tāze*	*boland*	*xošk*	*bozorg*	*ġermez, sorx*
MJS	*hāye*	*gerdu*	*gosne*	*tešne*	*tahl*	*tāze*	*derāz*	*hošk*	*gap*	*sohr*
SAS	*toxmemorġ*	*gerdu*	*gosne*	*tešne*	*tahl*	*tāze*	*bolond*	*hošk*	*gahp*	*sohr*
DNO	*hāga*	*gerdu*	*gosne*	*tešne*	*tahl*	*tāze*	*boland*	*hošk*	*gahp*	*sohr*
SEP	*toxm*	*gerdu*	*gosne*	*tešne*	*tahl*	*tāze*	*derāz, boland*	*hošk*	*gahp*	*sohr*
LGI	*xāve, xāye*	*gerdu*	*gosne*	*tešne*	*tahl*	*tāze*	*boland, derāz*	*hošk*	*gahp, dalu*	*sohr*
FRS	*tox, toxmorġ*	*gerdu*	*gošna*	*tešne*	*tahl*	*tāze*	*derāz*	*hošk*	*gahp*	*sohr*
JNB	*toxmemorġ*	*gerdu*	*gošne*	*tešne*	*tahl*	*tāze*	*derāz*	*hošk*	*gap, gahp*	*ġermez*
ARD	*toxm, tox*	*gerdu*	*gosne*	*tešne*	*tahl*	*tāze*	*derāz*	*hošk*	*gahp*	*sohr*
SHL	*toxmorġ*	*gerdu*	*gošne*	*tešne*	*talx*	*tāze*	*derāz*	*xošk*	*gahp, gonde*	*sorx*
BLB	*toxmemorġ, toxm*	*gerdu*	*gošne*	*tešne*	*talx*	*tāza*	*derāz*	*xošk*	*gonda*	*sorx*
LDG	*xāg, toxmorġ*	*gerdu*	*gosne*	*tešne*	*tahl*	*tāze*	*derāz, boland*	*xošk*	*gapu*	*sohr*
CHT	*tohmorg*	*gerdu*	*gosna*	*tešna*	*tahl*	*tāze*	*derāz*	*xošk*	*gapu*	*sohr*
CHC	*toxmemorġ*	*gerdu*	*gošne*	*tešne*	*talx*	*tāze*	*derāz*	*xošk*	*gonde, bozorg*	*sorx*
SKC	*toxmorġ*	*gerdu*	*gošne*	*tešne*	*talx*	*tāze*	*derāz*	*xošk*	*gonde, bozorg*	*sorx*
SHS	*toxmorġ*	*gerdu*	*gošne*	*tešne*	*talx*	*tāze*	*derāz*	*hošk*	*gahp, gonde*	*ġermez, sorx*
ARJ	*toxmorġ*	*gerdu*	*gošne*	*tešne*	*talx*	*tāze*	*derāz*	*xošk*	*gonde*	*sorx*
FTA	*toxmorġ*	*gerdu*	*gošne*	*tešne*	*talx*	*tāze*	*derāz*	*xošk*	*gonde*	*ġermez, sorx*
HAR	*toxmorġ*	*gerdu*	*gošne*	*tešne*	*talx*	*tāze*	*boland*	*xošk*	*bozorg, gonde*	*ġermez, sorx*
SKO	*toxmorġ*	*gerdu*	*gošne*	*tešne*	*talx*	*tāze*	*derāz*	*xošk*	*gonde, bozorg*	*sorx*
FSH	*toxmoġ*	*gerdu*	*gošne*	*tešne*	*talx, zaġġove*	*tāze*	*derāz*	*xoš*	*gonde*	*sox*
HAF	*toxmomorġ*	*gerdu*	*gošne*	*tešne*	*talx*	*tāze*	*derāz*	*xoš*	*gap, bozorg*	*sorx*
BOR	*toxmorġ*	*gerdu*	*gošne*	*tešne*	*talx*	*tāze*	*derāz*	*xošk*	*gonde*	*sorx*
NQC	*toxmorġ*	*gerdu*	*gošne*	*tešne*	*talx*	*tāze*	*derāz*	*xoš*	*gonde, bozorg*	*sorx*
BEN	*numurtā*	*ġoz*	*ājˇ*	*susuz*	*ājje*	*tazza*	*uzun*	*ġura*	*böyök, yekka*	*ġirmiza*
MAM	*nomurta*	*ġoz*	*ājˇ*	*susuz*	*ājje, ajje*	*tāzza*	*uzun*	*ġurre*	*yekka, beyig*	*ġermez, guli*
SKT	*numurta*	*ġoz*	*ājˇ*	*susuz*	*ājje*	*tāza*	*uzun*	*ġurre*	*böhig*	*ġermez*
KIN	*numurta*	*ġoz*	*ājˇ*	*susuz*	*talx*	*tāzza*	*uzun*	*ġurri*	*yekka, böhög, gonda*	*ġermez*
JNT	*numurta*	*ġoz*	*ājˇ*	*susuz*	*āči*	*tāza*	*uzun*	*ġurri*	*beyig*	*ġermez*
BLT	*numurta*	*ġoz*	*ājˇ*	*susuz*	*ājji*	*tāzza*	*uzān*	*ġurri*	*beyig*	*ġermez*
NQT	*numurtā*	*ġoz*	*ājˇ*	*suyoz*	*āčče*	*tāzze*	*uzun*	*ġurre*	*beyig*	*ġermez*
SUL	*yumurtā, numurtā*	*ġoz*	*ājˇ*	*suyuz*	*āčče*	*tāzzā*	*uzun*	*ġurre*	*beyig*	*ġermez*

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

variety	61. 'white'	62. 's/he came'	63. 's/he fell'	64. 's/he slept'	65. 's/he ate'	66. 's/he knew'	67. 's/he thought'	68. 's/he closed'	69. 's/he tied'
PRS	*sefid*	*āmad*	*oftād*	*xābid*	*xord*	*midānest*	*fekr kard*	*bast*	*bast*
MJS	*esbeyd*	*ovayd*	*vast*	*xavsid, xavsest*	*xard*	*dovnes*	*ferg kerd*	*bast*	*gerey zayd*
SAS	*esbir*	*aye*	*vast*	*xāfti*	*xard*	*davest*	*ferg kerd*	*bast*	*bast*
DNO	*esbid*	*avod*	*vast*	*xavsid*	*xard*	*dovest*	*diyār god, ferg kerd*	*bast, čeft kerd, zeyd*	*gere dād*
SEP	*esbid*	*oveyd*	*vast*	*xavsid*	*xard*	*dōnest*	*fek kerd*	*bast*	*bast*
LGI	*sebeyd*	*vey*	*vahs*	*xavsid*	*xa*	*davnes*	*fek kerd*	*bas*	*bas*
FRS	*esbi*	*uma, umad*	*vast*	*xavsid*	*xurd*	*idunest*	*fek kerd*	*bast*	*bast*
JNB	*espi*	*ovey*	*vast*	*xavsi*	*xord*	*idunest*	*fek kerd*	*bast*	*bast*
ARD	*esbid*	*avey*	*vast*	*xavsi*	*xard*	*dunest*	*fek kerd*	*bast*	*bast*
SHL	*esbid*	*oma*	*oftā*	*xavsid*	*xord*	*idunest*	*fek kerd*	*bast*	*bast*
BLB	*sefid*	*eveyd*	*oftād*	*xavsid*	*xord*	*idunest*	*fek kerd*	*bast*	*bast*
LDG	*sefid, espid*	*omey*	*vahs*	*xovsid*	*xard*	*dunes, idunes*	*ferk kerd*	*bas*	*bas, gere zeyd*
CHT	*safid*	*uma*	*vahs*	*xavsi*	*xa*	*dunes*	*fek ke*	*bahs*	*gere zey*
CHC	*sifid*	*umad*	*oftād*	*xābid*	*xord*	*balad bud, balad bid*	*fek kard*	*bast, piš kard*	*bast, gere zad, gere kard*
SKC	*sifid*	*umad*	*oftād*	*xābid*	*xord*	*midunest, balad bid*	*fek kard*	*bast*	*bast, gere zad*
SHS	*ispid*	*umad*	*oftād*	*xavsit*	*xord*	*dunest*	*fek kerd*	*bast*	*bast*
ARJ	*sifid, espit*	*umad*	*oftād*	*xābid, xovsid*	*xord*	*balad bud*	*fek kard*	*bast*	*bast, gere zad*
FTA	*sifid*	*umad*	*oftād*	*xābid*	*xord*	*idunest*	*fek kerd*	*bast*	*bast, gere zad*
HAR	*sefid*	*eveyd, umad*	*oftād*	*xābid*	*xord*	*idunest*	*fek kerd*	*bast*	*bast, gere zad*
SKO	*sifid, sefid*	*umad*	*oftād*	*xābid*	*xord*	*midunes*	*fek kard*	*bast*	*bast, gerov kard*
FSH	*sifid*	*umed*	*oftād*	*xābid*	*xord*	*balad bid, midunes*	*fek kard*	*bast*	*bast*
HAF	*sifid*	*umad*	*oftād*	*xābid*	*xord*	*idunest, balad bid*	*fek kerd*	*bast*	*gere zad*
BOR	*sifid*	*umad*	*oftād*	*xābid, xavsid, kape kard*	*xord*	*balede*	*fek kard*	*bast*	*bast*
NQC	*sifid*	*umad*	*oftād*	*xābid*	*xord*	*dunes*	*gemun kard*	*bast*	*gerey zad*
BEN	*āġ*	*galda*	*düšda*	*yātta*	*yeda*	*bilirda*	*fikr ilada*	*ürtde*	*bāġlāda*
MAM	*āġ*	*galde*	*dušde*	*yātde*	*yee*	*bilirde*	*fekr elir*	*bāġlae*	*bāġlae*
SKT	*āġ*	*galde*	*düšde*	*yātde*	*yede*	*bilir*	*fikr ilir*	*bāġlade*	*bāġlade*
KIN	*āġ*	*galdi*	*dišdi*	*yātdi*	*yedi*	*bildi*	*fekr etti*	*bāġladi, bāġlādi*	*bāġladi, bāġlādi*
JNT	*āġ*	*galdi*	*dušdi*	*yātdi*	*yeyi*	*beldi, beliyerdi*	*fekr eddi, fegr eddi*	*bāġlayi*	*bāġlayi*
BLT	*āġ*	*galdi*	*dišdi*	*yāddi*	*yedi*	*bildi*	*fekr eladi*	*bāġlādi*	*bāġlādi*
NQT	*āġ*	*galle*	*dišde*	*yādde*	*yede*	*bille*	*fikr edde*	*bāġlāde*	*bāġlāde*
SUL	*āġ*	*galle*	*dišde*	*yādde*	*yede*	*bilirde*	*fikr yedde*	*bāġlāde*	*bāġlāde*

variety	70. ‘s/he hit’	71. ‘it burned’	72. ‘s/he swept’	73. ‘here’	74. ‘there’	75. ‘under’	76. ‘today’	77. ‘yesterday’	78. ‘day before yesterday’
PRS	*zad*	*suxt*	*jāru kard*	*inǰā*	*ānǰā*	*zir*	*emruz*	*diruz*	*pariruz*
MJS	*zeyd*	*sohd*	*roft*	*ičo*	*očo*	*zeyr*	*amru*	*duš*	*parey*
SAS	*zeyr*	*soh*	*roft, jāru kerd*	*ičo*	*očo*	*zir*	*emru*	*duš*	*parir*
DNO	*zeyd*	*sohd*	*roft, jāru kerd*	*ičo*	*očo*	*zir*	*amru*	*duš*	*parir*
SEP	*zeyd*	*sohd*	*roft*	*ičo*	*ōčo*	*zēr*	*amru*	*duš*	*parir*
LGI	*zay*	*soh*	*roh, jāru kerd*	*ičo*	*učo*	*zeyr*	*amruz*	*duš*	*pariruz*
FRS	*zad*	*soh*	*ruft, jāru kerd*	*ičo*	*učo*	*zir*	*amru*	*duš*	*periru*
JNB	*zey*	*sohd*	*jāru kerd*	*ičo*	*učo*	*pāyn, zir, duman*	*amruz*	*duš*	*pariruz*
ARD	*zeyd*	*sohd*	*roft, jāru kerd*	*ičo*	*učo*	*zir*	*amruz*	*duš*	*periruz*
SHL	*za*	*soxt*	*roft, jāru kerd*	*inǰo*	*unǰo*	*zir*	*amruz*	*diriz*	*paririz*
BLB	*zeyd*	*soxt*	*jāru kerd, roft*	*inǰo*	*unǰo*	*zir*	*amruz*	*diruz*	*periruz*
LDG	*zeyd*	*soxt*	*roft*	*iro*	*uro*	*zir*	*emru*	*duš*	*pariru*
CHT	*zey*	*soht*	*roft, jāru ke*	*iro*	*uro*	*zeyr*	*omru*	*duš*	*parig*
CHC	*zad*	*suxt*	*jāru kard*	*inǰā*	*unǰā*	*zir*	*emruz*	*diriz*	*periruz, perruz, pariruz*
SKC	*zad*	*suxt*	*jāru zad, ruft*	*inǰā*	*unǰā*	*zir*	*amruz*	*diruz*	*peyruz*
SHS	*zad*	*soxt*	*roft*	*ičo*	*unǰa*	*zir*	*emruz*	*diriz*	*paririz*
ARJ	*zad*	*soxt*	*jāru kard, ruft*	*inǰā*	*unǰā*	*zir*	*amruz*	*diriz*	*peririz*
FTA	*zad*	*soxt*	*jāru kerd*	*ičo*	*učo*	*zir*	*amruz*	*diriz*	*peririz*
HAR	*zad, zeyd*	*suxt*	*jāru kerd, ruft*	*inǰo*	*unǰo*	*zir*	*amruz*	*diriz*	*peririz*
SKO	*zad*	*suxt*	*jāru kard*	*inǰā*	*unǰā*	*zir*	*amruz*	*diruz, dürüz*	*peyruz, payruz, pariruz*
FSH	*zad*	*sux*	*ruf*	*inǰā*	*unǰā*	*zir*	*amruz*	*dürüz*	*peyruz*
HAF	*zad*	*soxt*	*jāru kerd*	*inǰo*	*unǰo*	*zir*	*amruz*	*periz*	*peririz*
BOR	*zad*	*suxt*	*ruft*	*inǰā*	*unǰā*	*zir*	*emruz, amruz*	*diruz*	*peyruz*
NQC	*zad*	*sux*	*ruft*	*inǰā*	*unǰā*	*zir*	*amruz*	*diruz*	*peyruz*
BEN	*ʋurda, čālda*	*yānda*	*süpürda*	*burā*	*orā*	*ālt*	*bügün*	*dünan*	*iralegün*
MAM	*ʋurde, čālde*	*yānde*	*seperde, jāru elae*	*bure*	*ura*	*ālt*	*bigin*	*dineng, duneng*	*ilerigin*
SKT	*ʋurde*	*yānde*	*siperde*	*bura*	*ora*	*alt*	*begen*	*dünan*	*ilaregün*
KIN	*ʋurdi*	*yāndi*	*süpürdi, supurdi*	*burā*	*orā*	*āss, āssda*	*bügün*	*dünang*	*esrāǧigün*
JNT	*ʋurdi*	*yāndi*	*jāru eddi, seperdi*	*bureyi, bura*	*oreyi, ura*	*ālt*	*bigün*	*duneyi*	*esträǧigün*
BLT	*ʋurdi*	*yāndi*	*siperdi*	*burā*	*orā*	*ālt*	*begen*	*dinan*	*iraligin*
NQT	*ʋurde*	*yānne*	*jār edde*	*borā*	*orā*	*ālti*	*begin, begün*	*dünayn*	*ilaregin, ilaregün*
SUL	*ʋurde*	*yānne*	*sipirde*	*burā*	*orā*	*ālte*	*begin*	*dinayn*	*ilaregin*

variety	79. 'tomorrow'	80. 'day after tomorrow'
PRS	*fardā*	*pasfardā*
MJS	*sovah*	*pasovah*
SAS	*sove*	*pahsove*
DNO	*sovah*	*pahsovah*
SEP	*sovah*	*pahsovah*
LGI	*sohv*	*pasovā*
FRS	*sovā*	*passovā*
JNB	*sovā, sob*	*passovā*
ARD	*sov*	*pasovā*
SHL	*sovah*	*passovah*
BLB	*savb*	*passobā*
LDG	*sovā*	*passovā*
CHT	*sob*	*pansavā*
CHC	*sobā, fardā*	*passobā, pasfardā*
SKC	*sobā*	*passobā*
SHS	*sobā*	*passobā*
ARJ	*sobā*	*passovā*
FTA	*sobā*	*passobā*
HAR	*sobā*	*passobā*
SKO	*sobā*	*passobā*
FSH	*sovā*	*passobā*
HAF	*sobā*	*passobā*
BOR	*sobā, fardā*	*passobā*
NQC	*fardā, sobā*	*passobā, pasfardā*
BEN	*sahar*	*bürgün*
MAM	*sahar*	*birigin*
SKT	*sobā*	*birigün*
KIN	*sobā*	*passobā*
JNT	*sovā*	*passovā*
BLT	*sobā*	*passobā*
NQT	*sahar*	*birigin*
SUL	*sābā*	*birigin*

Appendix 2.
Inventory of sound correspondences identified in the lexical data

The following table lists the sound correspondences we identified in the lexical data. The clearest patterns are discussed in the article; these are presented first and listed according to the relevant article sections. Other correspondences are roughly grouped together according to theme and the order of items in the wordlist.

sound correspondence	item	examples	article section
ā/raised V before nasals	6. 'tongue'	*zabān/zebun*	4.2
	14. 'knee'	*zānu/zuni*	4.2
	16. 'bone'	*ostoxān/ostoxun*	4.2
	30. 'groom'	*dāmād/dumād*	4.2
	48. 'winter'	*zemestān/zemestun*	4.2
	49. 'house'	*xāne/hune*	4.2
	62. 's/he came'	*āmad/umad*	4.2
	66. 's/he knew'	*midānest/idunest*	4.2
	74. 'there'	*ānǰā/unǰā*	4.2
final vowel *a*/back V	5. 'mouth'	*dahan/dohun*	4.2
u/front V	1. 'hair'	*mu/mi*	4.3
	7. 'throat'	*gelu/geli*	4.3
	14. 'knee'	*zānu/zuni*	4.3
	17. 'blood'	*xun/xin*	4.3
	31. 'bride'	*arus/āris*	4.3
	77. 'yesterday'	*diruz/diriz*	4.3
	78. 'day before yesterday'	*pariruz/paririz*	4.3
non-high/high front V	7. 'throat'	*gelu/gilu*	4.3
	11. 'stomach (belly)'	*šekam/šikam*	4.3
	20. 'liver'	*ǰegar/ǰigar*	4.3
	61. 'white'	*sefid/sifid*	4.3
b/*ʊ* in simple coda	43. 'sun'	*āftāb/aftaʊ*	4.3
	45. 'night'	*šab/šaʊ*	4.3
	47. 'water'	*āb/aʊ*	4.3
b/*ʊ* between vowels	6. 'tongue'	*zebun/zeʊun*	4.4
	49. 'fox'	*rubā/ruʊā*	4.4
	79. 'tomorrow'	*sobā/soʊā*	4.4
	80. 'day after tomorrow'	*passobā/passoʊā*	4.4
b/*ʊ* in coda cluster	19. 'heart (organ)'	*ġalb/qalʊ*	4.6
	33. 'cat'	*gorbe/gorʊe*	4.6
x/*h*	28. 'girl'	*doxtar/dohdar*	4.4
	49. 'house'	*xune/hune*	4.4
	51. 'egg (of chicken)'	*xāye/hāye*	4.4
	55. 'bitter'	*talx/tahl*	4.4
	58. 'dry'	*xošk/hošk*	4.4
	60. 'red'	*sorx/sohr*	4.4
	71. 'it burned'	*soxt/soht*	4.4
CVC/VCC/ØC onset	11. 'stomach (organ)'	*šekam/eškam/kom*	4.4
CVC/VCC onset	46. 'star'	*setāre/āstāre*	4.4
	61. 'white'	*sefid/esbid*	4.4

(*Continued*)

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

(*Continued*)

sound correspondence	item	examples	article section
a/*e* word-internal	12. ‘stomach (organ)’	*mahde/mehde*	–
	21. ‘man’	*mard/merd*	4.4
	25. ‘child’	*bače/beče*	4.4
	38. ‘tree’	*daraxt/deraxt*	–
	48. ‘winter’	*zemestun/zemastun*	–
	67. ‘s/he did’ (in ‘s/he thought’)	(*fek*) *kard*/(*fek*) *kerd*	4.4
	78. ‘day before yesterday’	*pariruz/periruz*	–
fricative-liquid metathesis	55. ‘bitter’	*talx/tahl*	4.4
	60. ‘red’	*sorx/sohr*	4.4
kr/metathesis/*k*	67. ‘s/he thought’	*fekr kard/ferk kerd/fek kerd*	–
xo/*xa*	65. ‘s/he ate’	*xord/xard*	4.4
āb/*āft*/*aʊs*	64. ‘s/he slept’	*xābid/xāfti/xaʊsid*	4.4
ad/*eyd*	70. ‘s/he hit’	*zad/zeyd*	4.4
*u*CC/*o*CC	71. ‘it burned’	*suxt/soxt*	4.4
	72. ‘s/he swept’	*ruft/roft*	4.4
t/*d*/Ø	28. ‘girl’	*doxtar/dohdar/doʊar*	–
final *t*/*d*/Ø	71. ‘it burned’	*soxt/sohd/sox*	4.4
final *t*/Ø	38. ‘tree’	*daraxt/darax*	–
final *d*/Ø	30. ‘groom’	*dumād/dumā*	4.4
final *z*/Ø	76. ‘today’	*amruz/amru*	4.4
final *b*/*ġ*/Ø	39. ‘wood’	*čub/čuġ/ču*	4.4
final *k*/Ø	58. ‘dry’	*xošk/xoš*	–
n/Ø in demonstrative	73. ‘here’	*inǰā/ičo, iro*	4.4
	74. ‘there’	*unǰā/učo, uro*	4.4
č/*ǰ* in demonstrative	73. ‘here’	*inǰā/ičo*	4.4
	74. ‘there’	*unǰā/učo*	4.4
r/*l*	42. ‘leaf’	*barg/balg*	4.4
nasal/*ʊ*	14. ‘knee’	*zānu/zoʊi*	4.6
	30. ‘groom’	*dumād/doʊā*	4.6
	62. ‘s/he came’	*umad/oʊeyd*	4.6
	66. ‘s/he knew’	*idunest/doʊest*	4.6
d/*r*	12. ‘stomach (organ)’	*mehde/mehre*	4.6
	61. ‘white’	*esbid/esbir*	4.6
	70. ‘s/he hit’	*zeyd/zeyr*	4.6
ʊ/*y*	43. ‘sun’	*aftaʊ/aftay*	4.6
	45. ‘night’	*šaʊ/šay*	4.6
	47. ‘water’	*aʊ/ay*	4.6
q/*ġ*	19. ‘heart (organ)’	*qalb/ġalb*	4.6
g/*y*	20. ‘liver’	*ǰigar/ǰiyar*	4.6
	32. ‘dog’	*sag/say*	4.6
	51. ‘egg (of chicken)’	*hāga/hāye*	–
š/*s*	10. ‘finger’	*angošt/angost*	4.6
ros/*š*/*s*	53. ‘hungry’	*gorosne/gošne/gosne*	4.6

(*Continued*)

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press

(*Continued*)

sound correspondence	item	examples	article section
a/*e* word-final	12. ‘stomach (organ)’	*mehda/mehde*	4.6
	18. ‘urine’	*mesta/meste*	4.6
	22. ‘husband’	*mira/mire*	4.6
	23. ‘woman’	*zina/zine*	4.6
	25. ‘child’	*bača/bače*	4.6
	33. ‘cat’	*gorba/gorbe*	4.6
	54. ‘thirsty’	*tešna/tešne*	4.6
o/*e*	16. ‘bone’	*ostoxun/estexun*	–
a/*e*/*o*	76. ‘today’	*amru/emruz/omru*	–
a/*ā*	31. ‘bride’	*arus/āris*	–
ā/*a*/*o*	43. ‘sun’	*āftav/aftav/oftav*	–
ā/*o*	73. ‘there’	*unjā/unjo*	–
	36. ‘fish’	*māhi/mohi*	–
ā/*u*/*o*/*e*/*a*/Ø	62. ‘s/he came’	*āmad/umad/omey/eveyd/avey/vey*	–
ey/*i*	22. ‘husband’	*meyre/mire*	–
	23. ‘woman’	*zeyne/zine*	–
	61. ‘white’	*esbeyd/esbid*	–
	75. ‘under’	*zeyr/zir*	–
ā/*a*	35. ‘fox’	*ruvā/ruvah*	–
ā/*a*/*e*	80. ‘day after tomorrow’	*passobā/passovah/pahsove*	–
ey/*e*/*ov*	69. ‘s/he tied’	*gerey zayd/gere zad/gerov kard*	–
k/*g*	37. ‘scorpion’	*každin/gaždin*	–
ā/V*ž*	37. ‘scorpion’	*gādim/gaždim*	–
m/*n*	37. ‘scorpion’	*gādim/každin*	–
st/*s*(*s*)	9. ‘hand’	*dast/das*	–
	16. ‘bone’	*hast/xas*	–
	18. ‘urine’	*meste/mehse*	–
	48. ‘winter’	*zemestun/zemesun*	–
	63. ‘s/he fell’	*vast/vahs*	–
	66. ‘s/he knew’	*midunest/dunes*	–
	68. ‘s/he closed’	*bast/bas*	–

https://doi.org/10.1017/jlg.2021.8 Published online by Cambridge University Press