

INTRODUCTION

0.1. Overview of this dissertation

Burushaski is an isolated language spoken in northern Pakistan. There are a lot of languages from several language families and branches in the area, and the languages show some areal features. But studies on the languages have not been done well yet. In particular, among the other languages, Burushaski has no family language, so that we can neither substitute it with nor predict it from any other language for referring to it. This language would be required its own data in such as typological study.

The primary objective of this study is to provide a reference grammar of Burushaski (Hunza-Nager dialect) written in English, and to reexamine several points which previous studies have tended to merely reiterate without looking up and citing examples. My approach in this study is based on an analysis of my own database, which is developed through field research, and of previous research.

This dissertation is divided into two main parts (grammar and theoretical issues), two chapters (introduction (this chapter) and conclusions (chapter 12)), and two appendices (texts and vocabulary).

Part I (Grammar) includes chapters 1 to 8. Chapter 1 is for the phonological description. In chapter 2, I introduce preliminary information to describe and discuss grammar, including units such as words and clauses, word classes (a.k.a. parts of speech), and nominal classes (like genders in many languages). Chapters 3 to 7 chiefly deal with the morphology of word formation and derivation. Chapter 3 is for morphology of nominals, and chapter 4 is an analysis of pronouns and demonstrative and interrogative adjectives. Normal adjectives are described in chapter 5 with numerals, which behave more like nominals than verbals in Burushaski. Chapter 6 is devoted to verbal morphology starting from internal stem derivation, then continuing to conjugation and external deverbal derivation. The last chapter of morphology is chapter 7 where the other derivational morphological processes are explained: compounding, simple reduplication, echo-formation or fixed segment reduplication, and onomatopoeia and expressive formation. Chapter 8 deals with the syntax of Burushaski. It begins with basic constituent order in phrases and clauses; Burushaski is a typical head-final language so that modifiers basically precede the head noun and arguments are stated before the head predicate. Grammatical relations and information structure are also treated in this chapter as well as morphosyntactic description of several kinds of clauses

and reference.

Part II (Theoretical issues) consists of three chapters, 9 to 11. Chapter 9 “Transitivity and Its surroundings” is concerned with splits within the cognate stems of the same transitivity. There are dozens of verbal roots which have two stems of the same transitivity formed in different ways at the personal prefix; in this chapter I explore what motivates this, and ascertain that the motivation of split intransitivity is volitionality and split transitivity is caused by the likelihood of objects. Chapter 10 “d- Derivation” also covers verbal derivation. Here I investigate the function of a mysterious prefix *d-*, which has been problematic for previous scholars. The prefix functions as resultative, venitive, or anticausative according to the characteristics of verbal bases. Finally, chapter 11 “Definiteness and specificity” deals with nominal suffixes *-an* for singular and *-ik* for plural likely to be indefinite markers and a few morphosyntactic phenomena concerning definiteness, specificity, or referentiality. Such features may influence the choice of construction and/or stem types, and constituent order in clauses would be affected by informational importance.

Appendix I (Texts) includes four texts from my collection: *čhúmoe minás* (‘Fish tale’), *The Story of Hopar*, *šon gukúr*, and *kulió laskír*. *The Story of Hopar* is a historical tale in the Nager dialect about the origin of the Hopar residence. Two short tales of two shamans: *šon gukúr*, and of a witch: *kulió laskír*, and a long tale on a laughing fish: *čhúmoe minás*, are spoken in Hunza dialects, specifically, the Haiderabad (two short tales) and Ganish dialects (*čhúmoe minás*) (see Figure 4 in §0.2).

And Appendix II (Vocabulary) contains a list of about three thousand words. It consist of all the Burushaski words in this dissertation (including appendix texts) and the basic words I have collected in fieldwork thus far. Besides the meanings of the words translated into English, I also provide morphological annotation of plural forms for nouns or imperfect and conjunctive participle forms for verbs, derivational relations, and information on the origin or on some relevant form(s) where known.

0.2. Geographic background

Burushaski (ISO 693-3: bsk) is spoken by about 100,000 people in separate two areas in northern Pakistan. The major valleys of the eastern Burushaski spoken area are Hunza and Nager which belong to the Karakoram Mountains and to Hunza-Nager District of Gilgit-Baltistan (a federal capital territory of Pakistan; formerly known as the Northern Areas), on one hand, and the major valley of the western area is Yasin which belongs to the Hindukush Mountains and to Ghizer District of Gilgit-Baltistan, on the other hand, see Figure 1. Thus I call the Burushaski spoken on the Hunza and Nager

side “Eastern Burushaski”, and the one on the Yasin side “Western Burushaski” hereafter.



Figure 1. Large map of Burushaski spoken area

Besides these major valleys, minor groups of Burushaski speakers live in several other areas. For example, in Ishkoman Valley next to Yasin (where most people speak mainly Khowar), in Gojal (a.k.a. Upper Hunza) Valley (where Wakhi is predominant), and around Gilgit District between the two areas of Burushaski (where Shina is predominantly spoken), see Figure 2. Munshi (2006) says that Burushaski has been spoken for over a century by about 300 people speak Burushaski in Srinagar the summer capital of Jammu and Kashmir, in the far northwest of India. In light of the grammatical characteristics of their Burushaski, it appears that these Burushaski speakers in Srinagar are the offspring of emigrants from the Nager valley.

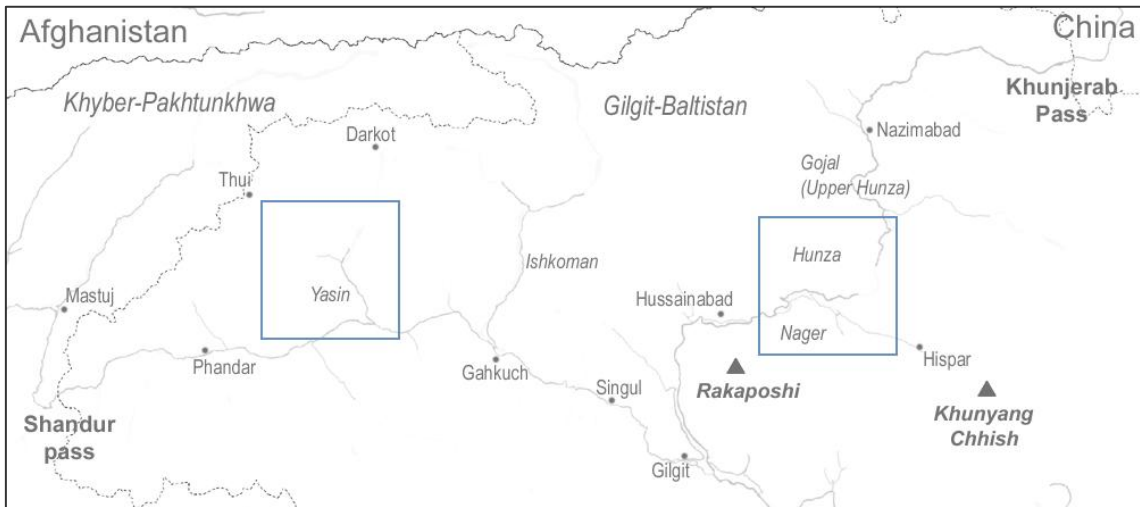


Figure 2. Western and Eastern Burushaski spoken areas



Figure 3. Yasin valley

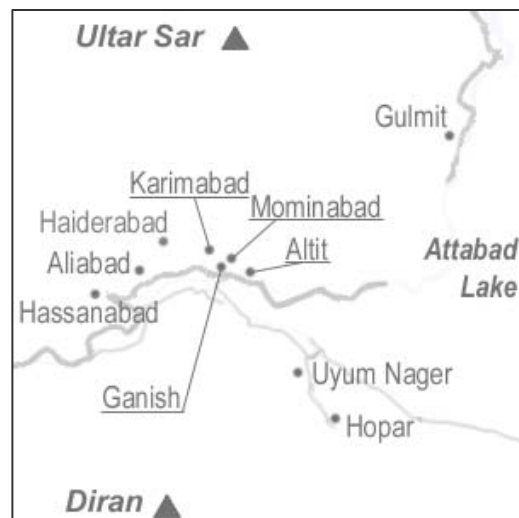


Figure 4. Hunza and Nager valleys

The Hunza and the Nager valleys face one another across the Hunza river, a tributary of the Indus (see Figure 4). Roughly speaking, the Hunza valley lies to the north of the river and the Nager valley to the south. Going east along the Hunza river, at the Ganish village of Hunza and the Sumiyar village of Nager, the Hunza river is joined by the Nager river coming from the southeast, and then both sides of the upper Hunza river, which turns north here, becomes the Gojal (or Upper Hunza) valley from the junction.

The nearest high peak from the settlement of the Hunza valley is Ultar Sar (7,388m) behind the town of Karimabad, Rakaposhi (7,788m) behind the Ghulmet village, Diran (7,266m) in the inner part of Hopar Valley, and Khunyang Chhish

(7,852m) behind the Hispar glacier are the nearest high peaks from the settlement of the Nager valley. The altitude of the residences of Burushaski speakers ranges about 1,000 up to over 3,000 meters. The terrain in Hunza and Nager is shaped by mountains belonging to subranges of the Karakoram Mountains, glaciers at the feet of mountains, and streams from the glaciers, so Burushaski speakers have settled along a radially branching rivers and streams.

Of Hunza and Nager, Hunza seems to be the main valley of Eastern Burushaski; in Nager, about 40 percent of the population speak Shina. Eastern Burushaski is directly bordered by the Gojal valley (in the Wakhi area, in the Wakhan Corridor of Afghanistan) to the north, the Shina area from the lower part of the Nager valley to around and south of the Gilgit city to the west. In a broader perspective, there are the Uyghur area, the Xinjiang Uyghur autonomous region beyond the Khunjerab pass to the northeast, the Balti area named Baltistan to the southeast, the Kashmiri area across Azad Kashmir of Pakistan to Jammu and Kashmir of India to the south, the Khowar wide area to the west beyond Western Burushaski and northwest which stretches to the Khyber-Pakhtunkhwa Province (formerly known as the North-West Frontier Province) of Pakistan.

In this way, Burushaski is situated at the intersection of these languages from different families, and this area is the north(west)most point of the Indian sprachbund. The surrounding languages are Wakhi (a Pamir language, Iranian, Indo-Iranian, Indo-European; spoken by Khik people), Shina (a Shina language, Dardic^{†1}, IE; by the Shin people), Khowar (a Chitrali language, Dardic, IE; by the Kho people), Uyghur (an Uyghuric language, Turkic), Balti (a Ladakhi language, Tibetan, Tibeto-Burman, Sino-Tibetan), and Kashmiri (a Kashmiri language, Dardic, IE). There are also small communities of speakers of two languages inside the Burushaski area, Domaaki (a Romani language, Central Indo-Aryan, II, IE; spoken by the Doma people) and Guj(a)ri (a Rajastani language, Western IA, II, IE; by the Gujur people) (see Figure 5).

^{†1} There are still disputes regarding the classification of Dardic, Western Indo-Aryan, and Domaaki. I treat the Dardic languages as a sub-group of the Indo-Iranian group alongside Indo-Aryan, Iranian, and Nuristani, unlike Morgenstierne (1973). And I classify the Western group of the Indo-Aryan separate from the Central group. Some linguists put Domaaki into the Dardic group as based only on geographic location and some features I consider just as areal. But I deny this assertion and classify the language as a Romani language owing to my own research on the language. Kausen (2006: 18) also annotates “oder ein Dialekt des zentralind. Domari?” tentatively placing the language into the Dardic group.

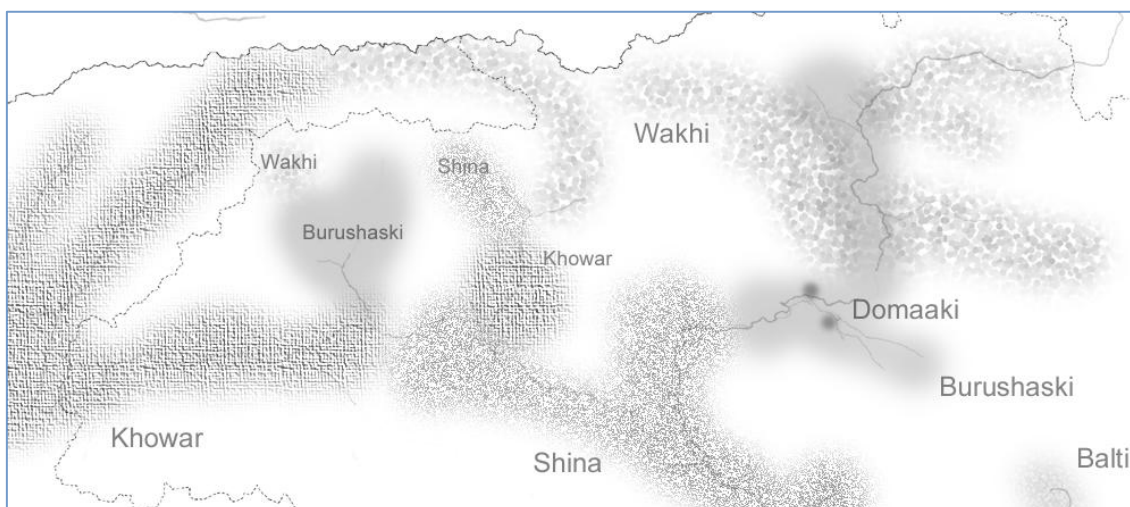


Figure 5. Map of Burushaski and the surrounding languages (based on Ethnologue (Lewis 2009) map^{†2})

0.3. Dialectology

All scholars of Burushaski unanimously agree that the most major dialectal divergence can be observed between Eastern (Hunza-Nager) Burushaski (EB) and Western (Yasin) Burushaski (WB). Comparing the 500-item basic vocabulary sets which I collected from three Eastern Burushaski consultants and one Western Burushaski consultant, 132 words (26%) of WB are not common to any of EB words. This numerical value verges upon Backstrom's (1992) result, see Table 1. And there are also many divergent grammatical features between Eastern and Western Burushaski. In phonology, Western Burushaski has no aspirated affricates and has lost the approximant with a retroflex feature (strictly it is not a retroflex approximant). In morphosyntax, it displays an optative forms of the copula for the first and the second person, the reduplicative imperfective stem formation, and a past predicative formation with a suffix *-asc* (employed for some nuance like background descriptions or topicalisation of the process of verb or mirativity (Lorimer 1935a: 436, 442; Berger 1974: 40–41; Grune 1998: 10–11; Tiffou 1999: 172; Bashir 2010: 14)), which is not observed in Eastern Burushaski. Western Burushaski has been considerably influenced by Khowar being the surrounding major language.

^{†2} The web page: “Languages of Pakistan: NORTHERN PAKISTAN”.
http://www.ethnologue.com/show_map.asp?name=PK&seq=10 (accessed 2012-01-17)

Table 1. Lexical similarity percentage (based on Backstrom 1992: 40)

Hussainabad				Hunza valley	Eastern Burushaski
96	Ganish				
91	95	Uyum Nager		Nager valley	
92	93	97	Hopar		
70	70	68	71	Yasin valley	Western Burushaski
67	69	67	70		

Eastern Burushaski can be classified into major and minor dialects. It can first be divided into the Hunza dialects and the Nager dialects. The Hunza dialects include the Hillside major group and the Riverfront minor group which shows more similarities with the Nager dialects. Ultimately separate dialects can be identified nearly down to the level of individual hamlets, villages, and towns based on details of vocabulary: for example, ‘egg’ is pronounced *tiján* in Hunza and *tigán* in Nager, but *tinán* only in the Ganish and Murtazabad villages (situated in the Hunza valley) irrespective of generation. The Nager dialect is influenced by Shina more than the Hunza dialect, and 40 percent of Nager people speak Shina as L1. A simplified view of dialectal diversity is shown in Figure 6.

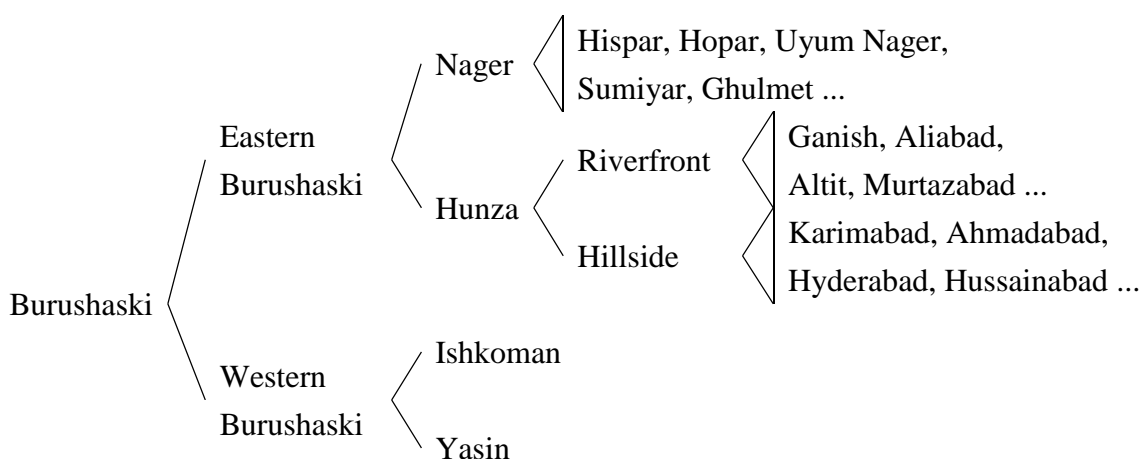


Figure 6. Dialects of Burushaski

0.4. Cultural background

Burushaski is originally the ethnic language of the Burusho people. They have lived here for over a thousand years. From where they came here to settle and when it was are not yet clear. The origin of the people is still as unknown as that of the language. DNA research by Wells et al. (2001) has tentatively grouped their ancestry with the

Bartangi (Afghani in Pamir region) and the Sinte Romani (Gypsy) peoples.

Burusho people are predominantly Muslims now, and further, almost all Hunza people are Isma‘ili (Nizari) and Nager people are Shi‘ite (Twelver), but the people of Ganish in Hunza belong to the Shi‘ite sect. The propagation of Islam in this area occurred quite early and there remains no religious vestige of the pre-Islamic age. There are only a few remains of prehistoric people in Hunza-Nager represented by the “sacred rocks”, on which there are many petroglyphs of human beings and mammals like ibexes and so forth carved in between 5000 BC to 1000 BC.

Up until 25th September, 1974 there were two states in this area for centuries, two independent principalities of Hunza and Nager. These princely states were governed by the Mirs, or as he is called, particularly in Hunza, *tham*, and formerly the Hunza valley was also governed by the Nager Mir. In 1947, the states of Hunza and Nager acceded to Pakistan but continued as semi autonomous states after that. And then in 1974 the government of Pakistan dissolved the Mirs of both states and these states were merged into the Northern Areas of Pakistan. The capital of the Hunza state was Baltit (today’s Karimabad) and the capital of Nager was Uyum Nager (cf. *uyúm* ‘big’). The states had much cultural exchange with Afghan, China, Kashmir and Tibet. There are a lot of loan words from Persian, particularly in royal vocabulary, which became a superstratum of Burushaski for a while because the Mirs could speak Persian.

People mainly earn a living with agriculture and traditional industries, but lifestyles are changing particularly among the people of Hunza, as they have also begun making a living with tourism and mountaineering since the Karakoram Highway has built and opened to the public in 1986. There is a remarkable economic gap between the valleys and only a few foreigners visit Nager. The gap has led to a visible difference of education, too, so that the literacy rate of both sexes in Hunza has risen to about 95% as the highest among all areas of Pakistan, while the rate in Nager has been at a lower level but it is said that the literacy rate in Nager is approaching to Hunza nowadays, provided that there are not indisputable statistical data on it and the numbers which exist are inconsistent data cited without a source. Now they are losing basic agricultural and other technical terms of Burushaski quickly, and getting a lot of new tools along with Urdu and English words.

0.5. Previous studies

There are two notable large studies on Burushaski: the first is Lorimer (1935–38) and the second is Berger (1998) both of which are comprehensive works in three volumes.

The first documentation of the Burushaski language is the work of Cunningham (1854), in which he recorded the vocabulary of ‘Khajunah’. This ‘Khajunah’ language seems equivalent to the Hunza-Nager Burushaski language. The name of “Burushaski” (actually “Boorishki”) can be seen in Biddulph (1880). He wrote a grammatical sketch of Burushaski covering a wide range of topics.

After such fragmentary documentations, D. L. R. Lorimer published the first descriptive grammar of Eastern Burushaski (with a grammatical sketch of Western Burushaski) in 1935 (vols. I and II) and 1938 (vol. III) which includes 47 texts and a list of about 7,500 or 8,000 words. He totally treated the grammar (vol. I), texts (vol. II; 67 texts), and vocabulary (vol. III; about 9,500 entries) of Eastern Burushaski with this work. After his work, Burushaski study seems to have become less popular once again.

Hermann Berger wrote a grammar of Western Burushaski (Werchikwār) in 1974 and after a quarter of century he published Berger (1998) which is the largest work on Eastern Burushaski up to now and which includes 67 texts and about 10,000 (sub)entries. Unlike Lorimer, Berger established the phonology before documenting the grammar, texts, and vocabulary.

Some studies deal in and deeply discuss individual grammatical issues: e.g., Tikkanen (1995) for converbs, Morin and Tiffou (1988) and Bashir (2004) for voice phenomena.

Burushaski stories have been well collected by Lorimer, Tikkanen (with English analyses), Berger, van Skyhawk (with German analyses), Tiffou (with French analyses), and so on. And Tiffou (1993) has recorded a great number of Burushaski proverbs.

There are several attempts to relate Burushaski to languages such as Basque, Yeniseian, Caucasian (or Dené-Caucasian), Dravidian, and Shino-Tibetan. Lately, on the genealogy of Burushaski, Čašule (1998, 2010, et al.) advocates that Burushaski is an Indo-European language. But his opinion has not had the backing of the other researchers yet but de la Fuente (2006).

Tiffou (2004b) summarizes the history and details of the other main previous studies on Burushaski well and further.

Among studies of the Burushaski, some works by Burushaski native speakers can be seen such as Hunzai (1984, 1998, 2003 among others) from Hunza, Shafi (2006) from Yasin, and Munshi (2006) from Srinagar. These studies include somewhat valuable information, are but seemingly highly subjective so that they make it hard to understand the actual situation of Burushaski, at least for non-native, readers. Hunzai and his Burushaski Research Academy are using and trying to spread the original Burushaski writing system adapted from Urdu or some other cognate ones; the writing system

includes several problems and they have not succeeded in its propagation yet. Burushaski speakers now tend to use ad hoc writing systems based on the Roman alphabet on facebook and other social media. (There is still no unified system among scholars now. As to the notation of each scholar, see Table of notations at page xv above.)

The following table shows the texts which mainly I referred to in this dissertation.

Table 2. Main reference texts from previous studies

Text title	Source	Sentences	Words	Dialect
Story of the North Wind and the Sun	Lorimer (1927)	8	116	Hunza
The Frog as a Bride	Tikkanen (1991)	506	5,038	Hunza
<i>Urke Yat</i>	Hunzai (1998)	11	126	Hunza
<i>ŽAKÚNE MARÁQ</i> (41 texts)	Hunzai (1998) Berger (1998b)	5 811	121 20,049	Hunza
Shiri Badat (26 texts)	Willson (2002) Berger (1998b)	53 353	624 9,885	Hunza Nager
<i>Híspara Šajirá</i>	van Skyhawk (2006)	24	1,534	Nager

0.6. Fieldwork

This dissertation will discuss not only data from previous studies, but also my own field data. I have conducted fieldwork on Burushaski six times from 2004 to 2009; the times and places of field surveys were as shown in Table 3.

Table 3. Summary of fieldwork

Year	From	To	At (Hunza; Nager; (Yasin))
2004	12 th Nov.	17 th Nov.	Karimabad;
2005a	21 st Feb.	10 th Mar.	Karimabad, Aminabad;
2005b	19 th Aug.	4 th Sep.	Karimabad, Mominabad;
2007	8 th Aug.	11 th Sep.	Karimabad, Mominabad; Hopar; (Taus, Ghojalti)
2008	7 th Sep.	17 th Nov.	Karimabad, Altit, Ahmadabad; Hopar, Uyum Nager
2009	16 th Aug.	11 th Sep.	Karimabad; Hopar

From the first time, I have been collecting Burushaski words and eliciting grammatical information in every survey. And I have recorded several stories from the work of 2005b.

Information about my principal consultants is provided in Table 4. They all are male.^{†3} There are more people (including women) who gave me information on Burushaski and/or told me Burushaski stories, which have not yet been fully transcribed.

Table 4. Personal information of principal consultants

Name	Born in	Birthplace	Edu.	Occupation (at the time)	Other languages
Afraz ul-Lah Beg	1929	Haidderabad	9	ex-Army	UR, EN
Essa Karim	1974	Aminabad	BA	Guide, Chairperson	UR, EN, J
Alamgir Khan	1944	Aminabad	0	Receptionist	UR, EN
Ejaz ul-Lah Baig	1970	Karimabad	MA	Curator, Librarian	UR, PE, EN
Musa Baig	1979	Ganish	12	Hotel owner	UR, J, EN
Liaqat Hussain	1984	Ganish	BA	Hotel stuff	UR, EN
Muhammad Ali	1963	Ganish	0	Farmer	UR
Ainur Khayat	1973	Hopar	8	Hotel stuff, Farmer	UR, SH
Muhammad Abbas	1972	Hopar	8	Hotel-stuff, Guide	UR, EN

Field research was conducted by means of Urdu, and I built up my database chiefly from field data from elicitation and story collection.

In this dissertation, if an example sentence has no source information, the example is from own elicitation data. Source information is always cited for examples from previous studies (see Table 2) and the stories which I collected (Table 5 below).

^{†3} Roughly speaking, Islamic societies tend not to let women go outside to public space. It is a reason for inclining towards male consultants. But sometimes I have had opportunity to speak with women in Burushaski and at the time I felt no difficulty in speaking with them. I think thus there would be no variance of Burushaski between men's and women's speaking.

Table 5. Main reference texts from my field data

Text title	Source	Sentences	Words	Dialect
<i>šhon gukúr</i>	Afraz ul-Lah Beg	15	130	Hunza
<i>kulíó laskír</i>	Afraz ul-Lah Beg	26	198	Hunza
<i>čhúmoe minás</i>	Muhammad Ali	351	4,820	Hunza (Ganesh)
<i>uskó jótíšo urkái ke uyúm</i> <i>γuníqış qhúuq</i>	Musa Beg	51	1,128	Hunza (Ganesh)
<i>uyúm dayánum búšan</i>	Musa Beg	124	1,019	Hunza (Ganesh)
The story of Hopar	Muhammad Abbas	28	370	Nager

0.7. Typological overview

Burushaski shows plenty of agglutinative characteristics, and there are many kinds of both prefixes and suffixes. Basic constituent order in a clause is SV/AOV (§8.3). A modifier precedes a modified word and a relative clause tends to precede a relativized word, which almost always requires a distal demonstrative according to the relativizer to become a diptych construction as a whole (§8.8). The syllable structure is CCVCC, but both onset and coda CC clusters are observed at the word initial and final position, respectively (§1.2). The language has the distinctive pitch accent system (§1.3).

From a typological viewpoint, Burushaski has the following peculiarities. There are four nominal classes (like genders in other languages) in Burushaski and all nominals belong to a class (or more than one classes) that is, briefly speaking, determined by whether it is a human male, human female, concrete thing, or abstract notion (§2.3). It has several dozen plural suffixes for nouns and there is no clear rule predicting the suffixes from the nouns, their matches are individually determined lexically fixed (§3.2.2). Nominals take case suffixes to decline in some of the following cases: absolutive, ergative, genitive, essive, dative, ablative, and several locational cases (§3.5). Morphologically the case alignment is an ergative system splitting in temporality and person-number (§9.3), while syntactically a predicate agrees by a suffix in the person-number-class of a subject argument, not absolutive argument (§§6.4 and 8.4). Some nouns which inalienably possessed by someone must take a personal prefix (§3.4), and the same personal prefix is employed on verbs to corefer with undergoer arguments (§6.3.2).

This language has some Indian areal linguistic features: the opposition of retroflex and dental consonant series (§1.1.1), echo-formations (§7.3), expressives and onomatopoeia (§7.4), and so-called conjunctive participles (§8.9.3). But, on the other hand, Burushaski lacks some of the characteristics which are commonly observed in the

large part of other languages; e.g., neither double causative nor classifiers in Burushaski. And there are some features commonly seen among the languages in northern Pakistan. Like *-ek* in Shina and *-ek/-aka* in Domaaki and *čik* in Balti; Burushaski has the general singular marker *-an* for noun to indicate the referent is indefinite (§§3.3 and 11). Burushaski and Shina have the distinctive pitch accent system (§1.3) in common, which does not seem to be in the surrounding languages like Khowar and Wakhi. From my fieldwork, it remains unclear whether or not the pitch accent of Domaaki is distinctive. The vigesimal numeral system (§5.2) can be considered as an areal feature as well.