A Micro-Typological Study of Shina
A Hindu Kush Language Cluster

Nina Knobloch
A Micro-Typological Study of Shina

A Hindu Kush Language Cluster

Abstract

In this thesis, 9 Indo-Aryan languages which have previously been classified as Shina languages were analyzed. A cognate analysis of basic vocabulary was conducted, in order to explore the relatedness of the languages. Furthermore, a selection of phonological, morphological, syntactic, and lexical features was analyzed, in order to explore areal patterns among the languages. The data mainly consisted of first-hand data, which has been collected for the project "Language contact and relatedness in the Hindu Kush region", but even previous descriptions of the languages were used. The results primarily confirmed hypotheses about the relatedness of the Shina languages, and showed interesting areal patterns. The data also suggested that the Shina languages share many typical features with other Hindu Kush Indo-Aryan languages, such as SOV word order, the use of postpositions, sex based grammatical gender, and moderately complex to complex syllable structures. Other features, such as aspiration, retroflexion, and case alignment in noun phrases showed more variation and could certainly be relevant for future studies on these languages.

Keywords

Micro-typology, Indo-Aryan languages, Greater Hindu Kush, Shina

Sammanfattning


Nyckelord

Mikrotypologi, Indoariska språk, Greater Hindu Kush, Shina
## Contents

1. **Introduction** ............................................. 2
2. **Background** .......................................... 3
   2.1 Micro-Typology ..................................... 3
   2.2 The Greater Hindu Kush ............................................. 3
   2.3 The Shina Languages ............................................. 5
      2.3.1 Classification within the Indo-Aryan Branch ............... 5
      2.3.2 Linguistic Setting and Language Use ....................... 5
      2.3.3 Classification within the Shina Branch ................... 5
      2.3.4 Languages or Dialects? ................................. 7
2.4 **Summary** ........................................... 8
3. **Aim and Research Questions** ............................. 9
4. **Method** ................................................ 10
   4.1 Data .................................................. 10
      4.1.1 Field Data .......................................... 10
      4.1.2 Other Data Sources .................................... 10
   4.2 Sample ................................................ 11
      4.2.1 Gilgit Shina ........................................ 12
      4.2.2 Gurezi Shina ........................................ 12
      4.2.3 Brokskat ........................................... 12
      4.2.4 Kohistani Shina (Palasi) ............................ 13
      4.2.5 Ushojo ............................................. 13
      4.2.6 Kundal Shahi ....................................... 13
      4.2.7 Kalkoti ........................................... 13
      4.2.8 Sawi ............................................. 13
      4.2.9 Palula ............................................ 13
   4.3 Cognate Analysis ..................................... 14
   4.4 Feature Analysis ..................................... 14
      4.4.1 Phonology .......................................... 14
      4.4.2 Morphology ......................................... 16
      4.4.3 Syntax ............................................. 16
      4.4.4 Lexicon .......................................... 17
   4.5 Overall Comparison .................................. 18
   4.6 **Summary** ........................................... 20
5. **Results** ............................................... 21
   5.1 Cognate Analysis ..................................... 21
   5.2 Feature Analysis ..................................... 22
      5.2.1 Phonology .......................................... 22
      5.2.2 Morphology ......................................... 29
      5.2.3 Syntax ............................................. 32
      5.2.4 Lexicon .......................................... 35
      5.2.5 Summary .......................................... 40
6. **Discussion** ........................................... 42
   6.1 Discussion of Results .................................. 42
      6.1.1 Cognate Analysis ..................................... 42
      6.1.2 Feature Analysis .................................... 43
   6.2 Discussion of Method .................................. 47
7. **Conclusions** ........................................ 48
   7.1 Answers to Research Questions ....................... 48
   7.2 Future Studies ....................................... 49
Abbreviations

Glosses

1  st person
3  nd person
AG agentive
CP clitic postposition
DIST distal
ERG ergative
F feminine
LOC locative
M masculine
NOM nominative
OBL oblique
PFV perfective
PL plural
PROX proximate
PST past
Q question marker
REFL reflexive
REM remote
SG singular
? morpheme unknown or uncertain
/ alternative gloss

General Abbreviations

A subject of transitive verb
GHK Greater Hindu Kush
HKIA Hindu Kush Indo-Aryan
IA Indo-Aryan
O object of transitive verb
OIA Old Indo-Aryan
S single argument of intransitive verb

Presentation Conventions

Previous studies have shown that many Hindu Kush Indo-Aryan languages have lexical tone. This is also the case for the Shina languages, which typically have a two-way tonal contrast (Liljegren, 2017: 122-123). However, in the transcriptions of the examples in this study, tone is not taken into account, if not specifically indicated. Furthermore, most examples are not glossed completely - unknown or uncertain morphemes are marked with '?'. Also, a distinction between phonologically distinct morphemes and clitics has not been made in the glosses.
1 Introduction

The Greater Hindu Kush (GHK) is the multilingual and linguistically diverse area that includes north-eastern Afghanistan, northern Pakistan and northern India (Kashmir). A variety of language families can be found here, the largest one of them being the Indo Aryan languages (IA). The Hindu Kush Indo-Aryan languages (HKIA) have retained certain features that languages outside of the area have lost, and they are, for the most part, not very well described. That is also the case for the Shina languages, which constitute a subgroup of the HKIA languages that is spread out over the whole area.

The goal of this micro-typological study of Shina is to compare 9 Shina languages according to selected phonological, morphological, syntactic, and lexical features. In addition, a cognate analysis of basic vocabulary is conducted. The selected languages are Brokskat, Kalkoti, Kohistani Shina, Kundal Shahi, Gilgiti Shina, Gurezi Shina, Palula, Sawi, and Ushojo. This sample represents the geographical spread and genealogical diversity of the Shina languages, as it includes representatives from all over the Hindu Kush region, and from all genetic clusters that have been hypothesized in previous studies.

As there has never been a comparative study on the Shina languages before, the results are expected to give a first insight into similarities and differences between the different Shina languages, which can help with a more detailed classification in the future. Furthermore, the analysis might point out interesting features and areal patterns that could be the subject of future studies. However, as time and the amount of data is limited, the study is only of exploratory nature.
2 Background

In this section, the theoretical framework for the analysis is presented. First, the theoretical background for micro-typology is explained (2.1). In the second part, the geography and the linguistic setting of the area of interest, the Greater Hindu Kush, is described (2.2). In the following section, the classification of the Shina languages within the branch of the Indo-Aryan languages is discussed (2.3.1), followed by an outline of previous studies of the languages, as well as their current use (2.3.2). Then, hypotheses about the classification of the Shina languages within their branch are presented (2.3.3). Finally, the question whether the languages in this study are to be treated as separate languages or dialects of the same language is discussed (2.3.4).

2.1 Micro-Typology

Linguistic typology is commonly defined as the study of cross-linguistic variation and linguistic diversity (Daniel, 2010: 44-45). Typological studies can either focus on the distribution and variation of one specific feature or category, or consider many of them. Traditionally, these studies have been done on the macro-level, meaning that they had a broad perspective and used a large, world-wide sample. However, the study of a certain area or a group of closely related languages can also show interesting patterns, which is where the term micro-typology becomes important. Micro-typology differs from macro-typology in the sense that is has a restricted context regarding the geographical distribution or genealogical classification of the language varieties under study. The term is also closely connected to areal typology and dialectology, which usually have similar restrictions. Bisang (2004) argues that dialectology, contact linguistics, and linguistic typology in fact have more to do with each other than previously assumed. According to him, the study of the three fields could provide us with new perspectives on linguistic variation (2004: 11).

Dahl (2008) discusses the connection between areal linguistics and language typology. The two fields overlap in their interests, since typologically interesting features tend to accumulate in certain areas, and both fields came from the motivation that the generic family tree model was not enough to explain certain changes in languages. However, areal linguistics focuses on the similarities between neighbouring languages, especially when they do not have the same origin. The goal is to define a linguistic area, a Sprachbund. Areal typology, on the other hand, focuses on classifying languages based on grammatical and phonological features, and since these features tend to cluster in certain areas, the geographical distribution of features is studied, instead of focusing on the area and its characteristics (2008: 1456). Basically, both fields are approaches to the same topic, but from different sides: areal linguistics starts from geography, whereas areal typology starts with features.

Koptjevskaja-Tamm (2010) discusses the role of language contact for linguistic typology, and concludes that research that considers both fields can help us to explore the history of a language and its speakers, as well as to see language in a larger context. According to her, a microtypology should capture “both the essential linguistic similarities and differences among the language varieties under consideration, as well as the distribution of the types across the region” (2010: 584), in order to contribute to its areal-typological characterisation.

2.2 The Greater Hindu Kush

The Greater Hindu Kush is the area that includes parts of north-eastern Afghanistan, as well as northern Pakistan and northern India (Kashmir) (Liljegren, 2014: 133-134), as it can be seen in Figure 1. This mountainous area is characterized by a high degree of multilingualism and linguistic diversity. The largest language family in the area are Indo-Aryan languages (IA), followed by Iranian and Nuristani languages, all three of them being branches of the Indo-Iranian languages. Also, a few representatives

1 Jammu and Kashmir (sometimes only “Kashmir”) is divided into a Pakistani- and Indian-administered part along a so called “Line of control”.
of Tibeto-Burman languages and Turkic languages can be found, as well as the isolate Burushaski (Liljegren, 2017: 110). Many languages in the area have remained undocumented, or have only been described partially, which is mainly due to the region being rather isolated geographically, but also due to many religious and territorial conflicts which have caused a lot of unrest for the people, and which have made it difficult to access the area for research.

Figure 1: The Greater Hindu Kush. (Map data ©2019 Google, created: 8/4/2019)

Being left in relative isolation in the mountains, the IA languages of the Greater Hindu Kush have retained a range of features, such as certain consonant clusters, and a tripartite sibilant system with a contrast between dental, retroflex and palatal, which Indo-Aryan languages outside of the area have lost (Bashir, 2003: 822). According to Liljegren (2017: 148-149), these features also include "large consonant inventories, the (lexical) use of contrastive tone, tripartite pronominal case marking alignment, final-position question markers, a high frequency of left-branching subordinate constructions, and the presence of multi-degree and multi-dimensional deictic systems". Other features, such as the basic SOV word order, or the use of adpositions, are shared with IA languages outside the region (Masica, 2001: 240-243).

In previous studies, the term Dardic languages has often been used to denote this special group of IA languages (Morgenstierne, 1961). Strand (2001: 11) sees the term as "a linguistic baseless classification that should be discarded," but Bashir (2003: 821-822) argues that it simply is a convenient umbrella term denoting the rather isolated languages in the region. Liljegren (2014: 135) proposes to use the term "Hindu Kush Indo-Aryan" (HKIA) when referring to the group of languages that has previously been called "Dardic", without implicating any genealogical classification, but rather recognizing the traits that the languages in the region tend to share with each other, as opposed to other IA languages outside of the area.
2.3 The Shina Languages

2.3.1 Classification within the Indo-Aryan Branch

The Shina languages are a subgroup of the HKIA languages, besides Pashai, Kunar, Chitral, Kohistani, and Kashmiri (Bashir, 2003: 824-825, Liljegren, 2014: 135). They are spread out over the whole region, with speakers as far west as in the Kunar province in eastern Afghanistan, in and around Gilgit town in the Gilgit-Baltistan province Pakistan, and even Ladakh in the north-eastern part of Indian-held Kashmir. The exact number of speakers of all Shina-languages is difficult to determine, but according to Ethnologue it must be close to 950 000 (Simons and Fennig, 2019).

2.3.2 Linguistic Setting and Language Use

There has never been a comparative study of the Shina languages, but some of them have been described, or at least sketched, previously. Several languages, such as Gilgiti Shina (Degener and Zia, 2008), Gurezi Shina (Ahmed, 2019), Palula (Liljegren, 2016), and Kohistani Shina (Schmidt et al., 2008) have been analyzed well already. Other Shina languages have not been described with great detail, but there are at least grammar sketches worth mentioning. These languages include Sawi (Buddruss, 1967), Brokskat (Ramaswami, 1982; Sharma, 1998), Kundal Shahi (Baart and Rehman, 2005; Rehman, 2011), Ushojo (Decker, 1992), and Kalkoti (Liljegren, 2009, 2013).

For most Shina languages, there are no established writing systems, but for some varieties, attempts have been made to implement such. Palula, for instance, has a very well established writing system that is based on the Urdu script, and used in a pre-school in the village where the language is spoken (Liljegren, 2016: 36-37). An attempt to create a standardized writing for Gurezi Shina has been made by Ahmed (2015: 104-118), although it is not clear to which extent the script is used by the speakers. For Kohistani Shina, an attempt to establish a writing system has been made as well (Schmidt et al., 2008: 14). For Ushojo, there have also been recent endeavors to develop a writing system at the Forum of Language Initiatives in Pakistan, and an alphabet book and Ushojo-Urdu vocabulary have been produced in 2016 (Henrik Liljegren, personal communication, 17/05/2019).

Kohistani and Schmidt (2006: 140) discussed the possibility of Gilgiti Shina being a standard variety of Shina, and a base for a standardized writing system, but due to religious conflicts and the low lexical similarity with other Shina varieties, this proved to be very difficult.

Except for the varieties spoken in Kohistan, most Shina languages have a relatively low status in the areas where they are spoken, meaning that they are usually used at home or in the local communities, but not in public fora or when communicating with outsiders. In Kohistan, where Shina is spoken south-east of the Indus river, the varieties have a higher status, due to the low rate of bilingualism in the region (Schmidt et al., 2008: 12-13) Here, outsiders will use Shina as well, when communicating with locals. Still, most Shina languages are very well preserved and passed on to the next generation, except for Kundal Shahi, which is an endangered language, as its speakers are shifting to using Hindko instead (Baart and Rehman, 2005: 3).

Most Shina speakers are Muslims, except for the speakers of Brokskat in northern India, which are Buddhists (Radloff, 1992: 99).

2.3.3 Classification within the Shina Branch

As for most of the HKIA languages, the Shina languages have not been studied much previously, especially not in a comparative way, which makes it difficult to classify them, and to trace their origin. However, there have been attempts to do so, and some of the theories will be summarized below.

As illustrated in the Glottolog-database (Figure 2), the Shina languages (here called “Shinaic languages”) can be divided into three larger sub-groups: Kohistanic Shina, Shina (consisting of three more

---

2 Combined number of speakers of Brokskat (bkk), Kalkoti (xka), Kundal Shahi (shd), Palula (phl), Savi (sdg), Shina (scl), Shina Kohistani (plk) and Ushojo (ush).
dialect groups itself), and the western Shinaic languages. Kundal Shahi and Brokskat are classified as outliers, not belonging to any of the groups (Hammarström et al. 2019). This classification is based on a variety of previous descriptions, of which a selection is summarized in the following sections.

Radloff (1992: 103-113) recognizes four geographical clusters of Shina languages: A northern cluster, which mainly includes the area in and around Gilgit valley, an eastern cluster, including the varieties spoken in Astor, Baltistan and Dras, the “Diamer cluster”, consisting of Chilas and other western areas, and the “Kohistan cluster”, including the languages that are spoken in the valleys of Jalkot, Palas and Kolai, and on the eastern side of the Indus river.

Regarding the classification of Ushojo, Decker (1992: 70) points out that it shares large parts of its vocabulary with Kolai Shina and Palas Shina, two Kohistani Shina varieties, which indicates that the language most likely is a Shina descendant and most closely related to Kohistani Shina. This has also been confirmed by the informants’ self-reported history (1992: 67-68).

According to Strand (2001: 251), the IA languages of the Greater Hindu Kush originated from the
"Sanskrit speech of the early Hindo-Aryas", who migrated to the area around 3500 years ago. Their regional dialect developed into the HKIA languages as we know them today. In terms of the Shina languages in particular, he argues that there were two dialect centres: Chilas and Gilgit (Strand, 2001: 253). They developed from the speech of some Aryas who went north to follow the Indus river through Indus Kohistan. After almost two millennia of peace, a long period of war and conflict caused many Shina speakers to leave their original settlements, which is why we can find many of the Shina languages scattered in the region today. Strand (2001: 253) also shows a particular interest in displaced Shina-speaking communities in Swat, Dir Kohistan and Chitral, where Palula, Sawi and Kalkoti are spoken today.

Bashir (2003: 824-825) describes three groups of Shina languages: the Kohistan group (which is the equivalent to Radloff’s Diamer and Kohistan clusters), the Astor group (Radloff’s eastern cluster) and the Gilgit group (Radloff’s northern cluster). She also discusses Palula and Sawi as outliers that are closely related to each other, but, according to her, do not belong to any of the groups. Kalkoti is classified as a part of the Kohistan languages, not belonging to the Shina languages at all.

Regarding the classification of Kundal Shahi, Baart and Rehman (2005: 17) suggest that the language genetically is a Shina language, but shows strong influences from Kashmiri, Hindko and even Kohistani and other North-Western IA languages. The Kundal Shahi verb for ‘to be’, for example, show similarities with Kohistani languages such as Indus Kohistani, Gawri and Torwali, not with Shina. On the other hand, the verb for ‘to do/ make’ agrees with Shina (Baart and Rehman, 2005: 17).

Liljegren (2009) discusses the Shina outliers Sawi, Kalkoti, and Northern and Southern Palula and argues specifically that Kalkoti is also a part of this cluster, as opposed to Bashirs [824-825] bashir2003 dariastic classification. He reconstructed features of their common proto-language and hypothesized two different migration routes, one that connects Southern Palula (Ashreti) and Sawi to Chilas, and one that links Northern Palula (Biori) and Kalkoti with Tangir (2009: 57-58) - explaining how the Shina descendants came to appear there.

Even though there is a lot of uncertainty about the classification of the Shina languages, there appears to be consensus about Gilgiti Shina, and the languages/dialects surrounding it, being one cluster (Radloff, 1992; Bashir, 2003), as well as the languages that are spoken in and around the Chitral district (Palula, Sawi, Kalkoti) constituting another western cluster (Bashir, 2003; Liljegren, 2009, 2013). Brokskat is clearly an outlier, probably a descendant from Gilgiti Shina (Strand, 2001: 253), and Kohistani Shina and Ushojo are possibly related as well (Radloff, 1992; Decker, 1992; Bashir, 2003). Kundal Shahi was suggested to be an outlier of Shina (Baart and Rehman, 2005: 17), even though its exact classification remains unclear.

2.3.4 Languages or Dialects?

As previous descriptions of the Shina languages often have focused on only one language at the time, instead of comparing them to each other, it is difficult to tell how similar or different they in fact are. The question of whether two speech varieties can be called "languages" or "dialects of the same language" has been discussed among linguists for a long time, without ever finding a satisfying answer. In most countries, the political borders simply determine if the varieties in question are treated as two separate languages or two dialects of the same language, but in the case of the Shina languages this is not possible, as they are usually only spoken by small communities in countries with partially disputed borders. Also, the cultural and religious identity of the speakers has to be taken into account as well.

In previous publications, Shina has sometimes been treated as being one language with many rather different dialects, and sometimes they are all seen as separate languages. In reality, it is probably a mixture of both: some of the languages are (sometimes only partially) mutually intelligible and very similar to each other, and could therefore be seen as dialects of the same language, whereas others are rather different and should therefore be treated as separate languages, even though they all have a common origin.

1 Meaning the Kohistani languages, a branch of the HKIA languages, not the Shina Kohistani language.
For now, in order to ensure consistency, all varieties will be called "languages", without making any assumptions about how similar or different they are. The analysis will hopefully provide some clarity on this topic, even though this question will not be discussed any further in this thesis.

2.4 Summary

The Shina languages are a sub-group of the HKIA languages, with representatives all over the GHK. Most of them have not been described much in previous studies, but there are hypotheses about certain geographical and genealogical clusters. These include a north-eastern cluster, with the tibetanized outlier Brokskat, a Kohistan-cluster, and a western cluster. Kundal Shahi is difficult to assign to any of these clusters - it possibly is another outlier.
3 Aim and Research Questions

The aim of this study is to create a micro-typological profile of 9 Indo-Aryan languages that are classified as "Shina". By analyzing a selection of phonological, morphological, syntactic, and lexical features, as well conducting a cognate analysis, the following research questions are attempted to be answered:

1. How do the Shina languages differ from each other lexically? Which languages stand out in some way?
2. Do the results from the cognate analysis confirm earlier hypotheses about the relatedness of the Shina languages among each other?
3. Which features are shared by all Shina languages?
4. Which features display more variation among the Shina languages?
5. How do the Shina languages differ from each other structurally? Which languages stand out in some way?
6. What can the results of the structural analysis tell us about any areal patterns?
4 Method

In this section, the methodology, data sources and sampling methods of the study are explained. After a description of the data sources and the sample in 4.1 and 4.2, the procedure for the cognate analysis is described in 4.3. Then, an explanation of the analysis of different phonological, morphological, syntactic, and lexical features is given in 4.4. In 4.5, the method for the final comparisons of the languages is explained, focusing on their structural and lexical similarity.

4.1 Data

4.1.1 Field Data

The data for this study has been collected by Henrik Liljegren for the project Language contact and relatedness in the Hindu Kush Region between 2015 and 2018 in a series of multilingual workshops in Islamabad (Pakistan), Kabul (Afghanistan), Faizabad (Afghanistan), and Srinagar (India), as well as during additional individual sessions. The project itself includes 60 language varieties that are spoken in the GHK, with the goal to create a typological profile over the region. It is supported by the Swedish Research council and hosted by the Department of Linguistics at Stockholm University. The following data sets from the project were used for the analysis:

- A word list of basic vocabulary (40 lexical items), based on the Automated Similarity Judgement Program (Wichmann et al., 2016)
- A list of numerals (1 - 50, 60, 70, 80, 90, 100, 110, 120, 200, 1000)
- A list of 95 kinship terms
- A valency questionnaire with 87 sentences, mainly based on the Leipzig Valency Classes project (Hartmann et al., 2013)
- An adapted version of Wilkin’s demonstratives questionnaire (Wilkins, 1999)

For each language, one native speaker was consulted. Most of these data sets, except for the demonstratives questionnaire, were elicited by giving the participants simple written questionnaires in a meta language of their choice (English, Urdu or Pashto), which they were asked to translate (using a script of their choice, most of them chose an Urdu-based script, some chose Latin script). Then, they were also asked to read their answers out loud, while an audio or video recording was made. In addition, they were also involved in discussing and comparing the data from their own language to the other languages. Later, IPA transcriptions were made. The answers for the demonstratives questionnaire were elicited by pointing at a certain object (e.g., a book lying on the table in front of the participant and the interviewer), and then asking the informant to translate a specific sentence that described the situation (e.g., ‘I like ___ book.’).

4.1.2 Other Data Sources

In some cases, the field data was not enough for conducting a complete analysis of the feature in question, which is why other data sources had to be considered as well. Their quality and precision varies a lot, but as there is very little material about most of the languages under study, these were the only possible sources to consider. An overview of the materials that were used is given in Table 1.

*The grammatical description was published when a large part of the thesis already had been written, which is why it was only considered in some cases.*
Table 1: Other data sources that were used in the analysis.

<table>
<thead>
<tr>
<th>Language</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brokskat</td>
<td>Ramaswami (1982); Sharma (1998)</td>
</tr>
<tr>
<td>Kalkoti</td>
<td>Liljegren (2013)</td>
</tr>
<tr>
<td>Kohistani Shina</td>
<td>Schmidt et al. (2008)</td>
</tr>
<tr>
<td>Kundal Shahi</td>
<td>Baart and Rehman (2005)</td>
</tr>
<tr>
<td>Palula</td>
<td>Liljegren (2016)</td>
</tr>
<tr>
<td>Sawi</td>
<td>Buddrus (1967); Degener and Zia (2008); Radloff (1999); Ahmed (2016); Qadir and Anand (2018); Ahmed (2019)</td>
</tr>
<tr>
<td>Gilgiti Shina</td>
<td>Ahmed (2016); Ahmed (2019)</td>
</tr>
<tr>
<td>Gurezi Shina</td>
<td>Ahmed (2016); Qadir and Anand (2018);</td>
</tr>
<tr>
<td>Ushojo</td>
<td>Decker (1992)</td>
</tr>
</tbody>
</table>

4.2 Sample

The 9 languages that were analyzed in this study are Gilgiti Shina, Gurezi Shina, Kohistani Shina, Ushojo, Kundal Shahi, Palula, Sawi, Kalkoti and Brokskat, as these were the available data sets. They represent the geographical distribution (see Figure 3) and diversity of the Shina languages well, as they include languages from all over the Hindu Kush region, as well as representatives from all clusters that have been hypothesized in previous studies. The language code, exact location of the speaker, as well as time and place of the elicitation are given in Table 2.

Table 2: The sample.

<table>
<thead>
<tr>
<th>Language</th>
<th>Project code</th>
<th>Location</th>
<th>Elicitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brokskat</td>
<td>bkk</td>
<td>Batalik, Kargil, Ladakh, Jammu &amp; Kashmir, India</td>
<td>Kargil, May 2018</td>
</tr>
<tr>
<td>Kalkoti</td>
<td>xka</td>
<td>Kalkot, Sheringal, Upper Dir, Malakand, Khyber Pakhtunkhwa, Pakistan</td>
<td>Islamabad, Oct 2015</td>
</tr>
<tr>
<td>Kohistani Shina (Palasi)</td>
<td>plk</td>
<td>Palas, Lower Kohistan, Hazara, Khyber Pakhtunkhwa, Pakistan</td>
<td>Islamabad, Aug 2016</td>
</tr>
<tr>
<td>Palula</td>
<td>phl</td>
<td>Ashret, Chitral, Malakand, Khyber Pakhtunkhwa, Pakistan</td>
<td>Islamabad, Oct 2015</td>
</tr>
<tr>
<td>Sawi</td>
<td>sdg</td>
<td>Sau, Naray, Kunar, Afghanistan</td>
<td>Kabul, Apr 2017</td>
</tr>
<tr>
<td>Gurezi Shina</td>
<td>scl (i)</td>
<td>Gurez, Bandipora, Kashmir, Jammu &amp; Kashmir, India</td>
<td>Srinagar, Apr/May 2018</td>
</tr>
<tr>
<td>Ushojo</td>
<td>ush</td>
<td>Chail, Swat, Malakand, Khyber Pakhtunkhwa, Pakistan</td>
<td>Islamabad, Oct 2015</td>
</tr>
</tbody>
</table>

5ISO code, with additional country specification for India (i) and Pakistan (p), when necessary.
6Origin, and usually place of residence of the speaker.
4.2.1 Gilgiti Shina

Gilgiti Shina is the language that has supposedly been named most frequently in previous literature. It is mainly spoken in Gilgit city, Gilgit-Baltistan, Pakistan, and often referred to as just "Shina" or "Gilgiti". However, in order to avoid confusion, it will be called "Gilgiti Shina" in this study. Kohistani and Schmidt (2006) discussed the language use and living situation of Shina speakers in Gilgit, pointing out that it is mostly used at home or at the marketplace, but not in public fora. They also discussed the possibility of Gilgiti Shina being a standard variety of Shina and a possible base for a standardized writing system, but according to them this was difficult to implement, due to religious disagreements and low lexical similarity with other Shina languages (2006: 140). According to Kohistani and Schmidt (2006: 139), it is spoken by approximately 145,000 people.

4.2.2 Gurezi Shina

Gurezi Shina (alt. "Guresi") is a language that is similar to Gilgiti Shina, but belongs to another dialect group called Astori (Radloff, 1992: 98). It is spoken in the Gurez valley in the north of Indian-administered Kashmir, close to the line of control between India and Pakistan. Ahmed (2019: 5) estimates a number of approximately 25,000 native speakers of Shina in Gurez.

4.2.3 Brokskat

Brokskat is an outlier that originates from Gilgiti Shina (Strand, 2001: 253). It is spoken in the Indian-administered part of Kashmir, in the Kargil and Ladakh districts, and the number of speakers is estimated to 10,000, according to Ethnologue (Simons and Fennig, 2019). The language has been in contact with Ladakhi and other Tibeto-Burman languages for a long time, which is why it has been influenced significantly by those languages. It is not mutually intelligible with other Shina languages. (Bashir, 2003: 889).
4.2.4 Kohistani Shina (Palasi)

Kohistani Shina is spoken in the Kohistan district in the valleys of Jalkot, Palas, and Kolai. It has been mentioned by Radloff (1992: 113) as one of the main Shina clusters. The informant for the current study speaks the Palasi dialect. Kohistani and Schmidt (2006: 140-141) describe the language as having a rather high status in the region on the left side of the Indus river, where even outsiders will use the language when communicating with locals. The rate of bilingualism is very low, especially among women. According to Ethnologue, it is spoken by approximately 401,000 people (Simons and Fennig, 2019).

4.2.5 Ushojo

Ushojo is spoken by approximately 2,000 people in the area around the Chail and Bishgram valleys in Swat in north-western Pakistan (Decker, 1992: 66). It is most closely related to Kohistani Shina (Kolai) (Decker, 1992; Kohistani and Schmidt, 2006: 69-72).

4.2.6 Kundal Shahi

Kundal Shahi is an endangered Shina-language spoken only in the village of Kundal Shahi in Pakistan. The language is not mutually intelligible with any other Shina language, and it has been in close contact with other languages in the region, such as Kashmiri and especially Hindko (Baart and Rehman, 2005: 3). Nowadays, there are only approximately 500 speakers, and the language is not passed on to the next generation. Instead, the children in the village speak Hindko, and often only partially understand Kundal Shahi. Still, its speakers showed a positive attitude towards the preservation of their language (Baart and Rehman, 2005: 5, 8).

4.2.7 Kalkoti

Kalkoti is a western Shina language that is closely related to Palula and Sawi. It is spoken by approximately 6,000 people in Kalkot village in Dir Kohistan (Liljegren, 2013: 129). Strand (2001: 254) classifies it as a dispersed dialect of Shina. It has been in close contact with other IA languages, Gawri and other Kohistani varieties, and has been influenced significantly by them. However, it still is a Shina language in its essential parts (Liljegren, 2009, 2013).

4.2.8 Sawi

Sawi is another western Shina language, closely related to Palula and Kalkoti (Liljegren, 2009: 7). It is partially mutually intelligible with Palula, but only with great effort. Usually, speakers will switch to a lingua franca, such as Pashto (Naseem Haider, personal communication, 19/02/2019). Most of its speakers are displaced in Afghanistan and Pakistan, but previously the language was spoken in the village of Sau by the Kunar river in eastern Afghanistan (Buddruss, 1967: 7). It is uncertain to which extent the language is spoken in the village today, but Buddruss (1967: 7) estimated about 100 households speaking the language, by the time of his study. There is no direct contact with Palula (or any of the other Shina language) anymore, as the people in the village have been integrated into the surrounding Gawarbati speaking community, however, they are aware of the existence of Palula. Gawarbati is another HKIA language, and it has influenced Sawi significantly (Morgenstierne, 1941: 9).

4.2.9 Palula

Palula (alt. “Phalura”, particularly in older literature) has two dialects: Southern Palula, spoken in the Ashret valley in the Chitral district in Northern Pakistan, and northern Palula, spoken in the Biori valley not far away from there (Liljegren, 2016: 3-4). It is closely related to Sawi and Kalkoti (Liljegren, 2016: 18-19), and has a high status in the two villages where it is spoken. Liljegren (2016: 1) estimated a total number of 10,000 speakers. However, as soon as its speakers have to communicate with someone from
outside of the village, they will switch to other languages, e.g. Khowar. Within the villages, children usually learn Palula as their first language, with few exceptions (Liljegren, 2016: 23-25).

4.3 Cognate Analysis

The data for the cognate analysis consisted of the 40 word list, the basic numerals (1 - 10, 20), selected basic kinship terms, and 11 simple verbs that were extracted from the valency questionnaire.

The decision, whether or not certain words were considered cognates, was based on whether the data suggested any common historical sound changes, and whether there were any recognizable patterns of sound changes when comparing the words in the different languages. However, only clear examples were grouped together, as soon as there was uncertainty the words were ignored - otherwise, a thorough historical analysis would have been needed, which would have exceeded the scope of this study.

4.4 Feature Analysis

Matras (2007: 37) found that language contact typically leads to an enrichment of phonological systems, or in some cases also a simplification. Consonants are more likely to be affected (2007: 37), as is prosody (2007: 39). Features such as word order or bound morphology were found to be less influenced by areality than the phonological features (Matras, 2007: 60, 62) - however, they were still investigated in order to ensure the inclusion of a wide variety of features from many different linguistic domains. Koptjevskaja-Tamm and Liljegren (2017: 205) identified a variety of lexico-semantic phenomena that could be indicators of areality, one of them being "area-specific lexicalizations and a shared or similar-looking internal organization of certain semantic domains". The analysis of such structures can therefore be highly relevant for the investigation of areal patterns in languages. The choice of features that were analyzed was mainly guided by three factors:

1. The aim to represent a wide variety of features
2. The aim to include features that have been previously recognized as being important for the region
3. The limits that were set by the availability of data

While the approach followed in this study is mainly data driven and exploratory, the results of this structural analysis are expected to show indications of areal influence. The phonological features syllable structure, aspiration, and retroflexion, have previously been analyzed by Liljegren (2017) in a micro-typological study of the HKIA languages, and were found to show interesting patterns which distinguish them from other IA languages that are spoken outside of the area. An analysis of syntactic and morphological features was more difficult, as the data mainly consists of word lists, and only the valency- and demonstrative questionnaires provided data for analyses of e.g. agreement patterns and word order. Accordingly, basic word order and order of adposition and noun phrase were analyzed, as well as morphosyntactic alignment and grammatical gender. The analysis of the kinship terms was mainly based on the work of Nikolayeva (2014) - here a selection of criteria that could be obtained from the field data was chosen. Numeral bases and numeral composition have also been analyzed previously by Liljegren (2017), and were found to be very relevant for the characteristics of languages in the region, especially the high representation of vigesimality as opposed to decimal systems, which are more common world wide (2017: 143). 2- or 3-way contrasts in demonstratives were also found to be typical for the HKIA languages (Liljegren, 2017: 145), and as a questionnaire was available, this feature was investigated as well. A description of the individual features surveyed in the study follows.

4.4.1 Phonology

The phonological features that were analyzed were syllable structure, aspiration and retroflexion. Given that the field data mostly consists of basic vocabulary, and therefore is less likely to contain loan words,
it might be possible that new phonological structures, which other studies have recognized, do not appear in the data, or that the data contains traces of older structures, which have not been recognized in previous studies.

**Syllable Structure**

The analysis of the syllable structure was based on the WALS chapter 12A (Maddieson 2013). The author proposes three categories:

- **Simple syllable structure**, where (C)V-syllables are the only ones allowed.
- **Moderately complex syllable structure**, where modest extensions of simple syllables, such as one consonant after the vowel (e.g. CVC), are allowed. Two-consonant clusters in the onset are also allowed, but only if they are restricted to certain types of consonants in some way (e.g. CCV, CCVC).
- **Complex syllable structure**, where more combinations, such as two consonants before a vowel (if they are not restricted), two or more consonants after the vowel or even three consonants in the onset, are allowed.

For this analysis, the IPA transcriptions of the 40 word list, kinship terms, and numerals, as well as the corresponding audio files were used. Their syllable structures were annotated, in order to investigate which combinations were permitted.

**Aspiration**

According to Masica (1991:101), a four way contrast between voiced, voiced-aspirated, voiceless, and voiceless-aspirated, is typical among IA languages. However, among the IA languages in the GHK, a three-way contrast is more common, i.e. a contrast between voiceless, voiceless-aspirated, and voiced consonants (Liljegren 2017:120). For the analysis of the occurrence of aspiration contrast in the Shina languages, two criteria were investigated:

- Presence of contrast in voiced consonants
- Presence of contrast in voiceless consonants

Due to the limited number of data sets, other sources (besides the 40 list, kinship terms, and numerals) had to be considered as well, in order to discuss the phonemic status of the aspirated consonants.

**Retroflexion**

According to Liljegren (2017:116), retroflexion is a highly relevant feature for languages in the GHK. The following criteria were therefore investigated:

- Presence of retroflex plosives (/ʈ/, /ɖ/ in contrast with /t/, /d/) 
- Presence of retroflex fricatives (/ʂ/, /ʐ/ in contrast with /s/ and /ž/ or /ʃ/ and /ʒ/)
- Presence of other retroflex consonants (/ɳ/, /ɽ/, /ʈ͡ʂ/, /ɖ͡ʐ/)

Again, the data sources were the 40 list, the list of kinship terms, and the list of numerals. As for the analysis of aspiration contrast, other data sources had to be considered as well in order to make an attempt to prove the phonemic status of the consonants in question.
4.4.2 Morphology

Grammatical Gender

According to Corbett (2013a), the main criteria for the existence of grammatical gender is whether there are any other words in the sentence, that agree with the noun. If a language has grammatical gender, there is always a reflex of the gender system outside of the noun. Besides determining whether grammatical gender is present in the languages, it was analyzed whether the languages have a grammatical gender system that is sex-based, as in WALS chapter 31 (Corbett, 2013b), or animacy-based. The three values for this feature are therefore:

- Presence of grammatical gender
- Presence of sex-based grammatical gender
- Presence of animacy based grammatical gender

The data for this analysis consisted of the valency questionnaire, as more complex constructions were needed in order to observe agreement patterns.

Morphosyntactic Alignment

This feature concerns the alignment of case marking in noun phrases, as it is described in WALS chapter 98A (Comrie, 2013a). S is the single argument of an intransitive verb. A is the more agent-like argument of a transitive verb, whereas P is the more patient-like (Comrie, 2013a). There are two major types of alignment, nominative-accusative and ergative-absolutive. In a nominative-accusative system, A and S are marked in the same way (nominative), and P is marked differently (accusative). In an ergative-absolutive system, S and P are marked in the same way (absolutive), and A is marked differently (ergative). Languages can also have a tripartite system, where all three arguments are marked differently, or they have no morphological marking at all and establish the positions in other ways (e.g. word order). They can also have a split system, where certain constructions use ergative-absolutive marking and other nominative-accusative. The following values were set:

- Nominative-accusative
- Ergative-absolutive
- Tripartite
- Split ergative/accusative
- No case marking

The data source for this analysis was mainly the valency questionnaire, where the semantic roles agent and patient were assigned to the noun phrases, and then analyzed how they are marked morphologically. Even for this analysis, other data sources had to be considered, as it was difficult to find enough examples in the field data.

4.4.3 Syntax

Basic Word Order

For the analysis of the basic word order, i.e. the order of subject (S), direct object (O) and verb (V), the valency questionnaire was used as well. As it is described in WALS chapter 81A (Dryer, 2013b), there are 7 possible patterns:

- SOV
In order to analyze the word order patterns in the Shina languages, a variety of transitive sentences from the valency questionnaire were analyzed, and subject, verb, and object were identified.

**Adpositions**

Adpositions express the grammatical and semantic relation of the noun phrase (NP) with another unit in the clause. Their position in relation to the noun phrase has been analyzed, as in the WALS chapter 85A "Order of Adposition and Noun Phrase" (Dryer, 2013a). The categories are:

- Postpositions, where the adposition comes after the NP
- Prepositions, where the adpostion comes before the NP
- Inpositions, where the adposition is within the noun phrase
- No dominant order
- No adpositions

The valency questionnaire was the main data source for this analysis.

**4.4.4 Lexicon**

**Numeral Bases and Numeral Composition**

The list of numerals was used to identify the numeral bases. The following values were set, based on the WALS chapter 131 A (Comrie, 2013b):

- Decimal (base 10)
- Vigesimal (base 20)
- Hybrid decimal-vigesimal
- Other systems

However, it has to be noted that the classification of vigesimal and hybrid systems was done in a different way than in WALS. Comrie (2013b) considers systems that have vigesimal bases up to 99, and then switch to a decimal system from 100 onward, hybrid decimal-vigesimal systems. However, in the current analysis, languages that have vigesimal bases up to 99 were considered vigesimal systems, irrespective of whether the system switches afterwards, or not.

Furthermore, the numeral composition of the languages was analyzed, meaning the order of the base components 10 and 20, and the single digit (here denoted with $x$). This feature has been found to be a very relevant for the analysis of numeral systems in the languages in the GHK (Liljegren, 2017: 144). The following values were set for the analysis:
Kinship Terms

The strategy for the analysis of the semantic classification of kinship terms was mainly based on the work of Nikolayeva (2014: 79). The goal was to analyze, if the languages lexicalize certain contrasts, i.e. have a single lexeme that expresses the contrast, or if they need to use constructions of several lexemes on order to express the contrast. A simple example would be the words for grandparents in Swedish and German: in Swedish the words ‘mormor’ and ‘farmor’ indicate, whether the maternal or paternal grandmother is meant - the contrast is lexicalized, whereas in German, both of them would be denoted with ‘Großmutter’ and in order to express the contrast, additional lexemes would be needed. The following distinctions were analyzed, using the data from the kinship terms questionnaire:

- kinship line (‘mother’/’father’ vs. ‘parent’)
- direction of kinship (‘father’s father’ vs. ‘mother’s father’, ‘father’s mother’ vs. ‘mother’s mother’, ‘son’s son’ vs. ‘daughter’s son’ , ‘son’s daughter’ vs. ‘daughter’s daughter’)
- age (‘older brother’ vs. ‘younger brother’, ‘older sister’ vs. ‘younger sister’)

This analysis is by far no complete analysis of the semantics of kinship terms in the Shina languages, but it provides a brief insight into some distinctions that are made, and how the languages might differ from each other.

Demonstratives

According to Liljegren (2017: 145), a 2-way or 3-way contrast in demonstratives is typical for the HKIA languages. For the analysis of distance contrast and visibility contrast in Shina, the demonstratives questionnaire provides a variety of different situations where objects as close as on the speaker’s body, and as far as several kilometers away from the speaker, or even objects that were out of sight, were pointed out. The different situations and sentences from the questionnaire that were used in the analysis are given in Table 3. The goal was to identify how many distance- or visibility-based contrasts there were, and the following values were set:

- 2-way contrast
- 3-way (or more) contrast

4.5 Overall Comparison

Finally, the languages’ results of the cognate analysis and feature analysis were all compared to each other, in order to determine their lexical and structural similarity. For both comparisons, the software Splits Tree (Huson and Bryant, 2006) was used to visualize the results. First, a matrix with the results from the cognate analysis was created: all languages that share a cognate for a certain word, were assigned the same symbol in that column. From the resulting matrix, the software calculates a distance matrix that shows how many percent of the words in one language, are cognates with words in each of the other languages. The software can then generate a Neighbour Net, that shows how similar or different the languages are from each other. An example is shown in Figure 4. The length of the lines
Table 3: Elicitation sentences for analysis of demonstratives, based on Wilkins (1999).

<table>
<thead>
<tr>
<th>Situation</th>
<th>Sentence(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A ball hits the speaker on the finger.</td>
<td>‘___ ball hit me on the finger.’</td>
</tr>
<tr>
<td>A bug is sitting on the speaker’s arm.</td>
<td>‘___ bug is bothering me.’</td>
</tr>
<tr>
<td>A book lies on the table in front of them.</td>
<td>‘Is ___ your book?’</td>
</tr>
<tr>
<td></td>
<td>‘I like ___ book!’</td>
</tr>
<tr>
<td></td>
<td>‘Do you want to borrow ___ book?’</td>
</tr>
<tr>
<td>A ball lies on the ground.</td>
<td>‘Is that your ball?’</td>
</tr>
<tr>
<td></td>
<td>‘I would like to borrow ___ ball’</td>
</tr>
<tr>
<td>A house on the other side of the street.</td>
<td>‘___ house is beautiful!’</td>
</tr>
<tr>
<td></td>
<td>‘Do you know who lives in ___ house?’</td>
</tr>
<tr>
<td>A street behind the house.</td>
<td>‘___ street behind the house is very narrow.’</td>
</tr>
<tr>
<td></td>
<td>‘Do you know the number of ___ street?’</td>
</tr>
<tr>
<td>Mountains that are a few kilometers away.</td>
<td>‘Can you see ___ peak up in the hills?’</td>
</tr>
<tr>
<td></td>
<td>‘Have you been to ___ peak?’</td>
</tr>
<tr>
<td>A village behind the mountains.</td>
<td>‘Do you know ___ village behind the hills?’</td>
</tr>
<tr>
<td></td>
<td>‘Have you been to ___ village behind the hills?’</td>
</tr>
</tbody>
</table>

Figure 4: Example of SplitsTree representation of results (Created with SplitsTree 4; Huson and Bryant, 2006).

between the languages indicates their distance from each other, the longer the line, the fewer cognates they have.

For the analysis of the structural similarity, the strategy was similar to that of the lexical similarity. Here, the results from the feature analysis were represented in a matrix, where the languages, that share the same traits for a specific feature, were assigned the same number. A particular focus was put onto only considering binary values, with the two categories yes or no. For example for the analysis of grammatical gender, all languages that have grammatical gender would get the number 1, and all languages that do not have grammatical gender would get the number 0. The number (or symbol) used does not matter in the process, as long as the different groups each get a distinct symbol. There is no ranking between the symbols either. Then, a distance matrix was calculated again, and the distances were visualized in a Neighbour Net.
4.6 Summary

In this study, the following features were analyzed: Syllable structure, presence of aspirated consonants, presence of retroflex consonants, presence of grammatical gender and whether it is sex-based or not, morphosyntactic alignment in noun phrases, basic word order, order of adposition and noun phrase, numeral bases, numeral composition, semantic classification of kinship terms, and distance contrast in demonstratives. Furthermore, a cognate analysis was conducted. The lexical and structural similarity of the languages was analyzed by drawing a Neighbour Net with the software Splits Tree.
5 Results

In this section, the results of the cognate and feature analysis are presented. All examples are taken from the field data, if not explicitly marked.

5.1 Cognate Analysis

A cognate analysis was conducted for a list of 95 words in total. The resulting distance matrix, which illustrates the lexical similarity of the languages, is shown in Table 4. The languages are in general very similar to each other, varying between 51% (scl (i) - sdg) and 84% lexical similarity (scl (p) - plk).

Table 4: Distance matrix of cognate analysis.

<table>
<thead>
<tr>
<th></th>
<th>bkk</th>
<th>xka</th>
<th>plk</th>
<th>shd</th>
<th>phosph</th>
<th>sdg</th>
<th>scl (p)</th>
<th>scl (i)</th>
<th>ush</th>
</tr>
</thead>
<tbody>
<tr>
<td>bkk</td>
<td>0.327</td>
<td>0.360</td>
<td>0.379</td>
<td>0.304</td>
<td>0.415</td>
<td>0.36</td>
<td>0.429</td>
<td>0.24</td>
<td>0.281</td>
</tr>
<tr>
<td>xka</td>
<td>0</td>
<td>0.327</td>
<td>0.379</td>
<td>0.304</td>
<td>0.415</td>
<td>0.36</td>
<td>0.429</td>
<td>0.24</td>
<td>0.281</td>
</tr>
<tr>
<td>plk</td>
<td>0.237</td>
<td>0.360</td>
<td>0.360</td>
<td>0.304</td>
<td>0.415</td>
<td>0.36</td>
<td>0.429</td>
<td>0.24</td>
<td>0.281</td>
</tr>
<tr>
<td>shd</td>
<td>0.386</td>
<td>0.393</td>
<td>0.379</td>
<td>0</td>
<td>0.415</td>
<td>0.36</td>
<td>0.429</td>
<td>0.24</td>
<td>0.281</td>
</tr>
<tr>
<td>phosph</td>
<td>0.328</td>
<td>0.270</td>
<td>0.36</td>
<td>0.304</td>
<td>0.415</td>
<td>0</td>
<td>0.230</td>
<td>0.386</td>
<td>0.413</td>
</tr>
<tr>
<td>sdg</td>
<td>0.43</td>
<td>0.3</td>
<td>0.429</td>
<td>0.415</td>
<td>0.304</td>
<td>0</td>
<td>0.412</td>
<td>0.487</td>
<td>0.356</td>
</tr>
<tr>
<td>scl (p)</td>
<td>0.262</td>
<td>0.295</td>
<td>0.159</td>
<td>0.422</td>
<td>0.386</td>
<td>0.412</td>
<td>0</td>
<td>0.173</td>
<td>0.284</td>
</tr>
<tr>
<td>scl (i)</td>
<td>0.281</td>
<td>0.4</td>
<td>0.24</td>
<td>0.418</td>
<td>0.45</td>
<td>0.486</td>
<td>0.173</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>ush</td>
<td>0.461</td>
<td>0.413</td>
<td>0.338</td>
<td>0.464</td>
<td>0.383</td>
<td>0.356</td>
<td>0.284</td>
<td>0.3</td>
<td>0</td>
</tr>
</tbody>
</table>

Average: 0.339 0.354 0.312 0.397 0.339 0.387 0.299 0.354 0.379

Words that were especially likely to be cognates in all 9 languages were:

- Pronouns, e.g. ‘I’, ‘we’, and ‘you’
- Nature-related terms, e.g. ‘star’, ‘sun’, and ‘leaf’
- Body-part terms, e.g. ‘ear’, ‘eye’, ‘hand’, ‘tooth’ etc.
- Simple verbs, e.g. ‘go’, ‘give’, ‘eat’, ‘fear’
- Numbers
- “Simple” kinship terms such as ‘father’, ‘mother’, ‘brother’, ‘sister’ etc.

A small selection is shown in example 1.

(1) Examples for cognates in all 9 Shina languages.

<table>
<thead>
<tr>
<th></th>
<th>‘I’</th>
<th>‘star’</th>
<th>‘tooth’</th>
<th>‘go’</th>
<th>‘four’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brokskat</td>
<td>mo:</td>
<td>turi</td>
<td>dæni</td>
<td>gjæ</td>
<td>fœ:r</td>
</tr>
<tr>
<td>Kalkoti</td>
<td>ma tar</td>
<td>dæn</td>
<td>gi</td>
<td>fœ:r</td>
<td></td>
</tr>
<tr>
<td>Kohistani Shina</td>
<td>mõ: taru</td>
<td>don</td>
<td>gej</td>
<td>fœ:r</td>
<td></td>
</tr>
<tr>
<td>Kundal Shahi</td>
<td>mã tar</td>
<td>dand</td>
<td>gi</td>
<td>fœ:r</td>
<td></td>
</tr>
<tr>
<td>Palula</td>
<td>ma tōro</td>
<td>dānd</td>
<td>gi</td>
<td>fœ:r</td>
<td></td>
</tr>
<tr>
<td>Sawi</td>
<td>ma tari</td>
<td>dandu</td>
<td>gej</td>
<td>fœ:r</td>
<td></td>
</tr>
<tr>
<td>Gilgit Shina</td>
<td>ma tāro</td>
<td>don</td>
<td>gaj</td>
<td>fœ:r</td>
<td></td>
</tr>
<tr>
<td>Gurezi Shina</td>
<td>mu taru</td>
<td>donj</td>
<td>gej</td>
<td>fœ:r</td>
<td></td>
</tr>
<tr>
<td>Ushojo</td>
<td>ma: taru</td>
<td>dond</td>
<td>gaj</td>
<td>fœ:a</td>
<td></td>
</tr>
</tbody>
</table>

7 These percentages were calculated by subtracting the value for the distance between the two languages from 1. They represent portion of words that are cognates in the two languages under analysis.
Other words, including the more complex kinship terms, and individual words from other categories, showed more variation. A selection of examples is shown in (2).

(2) Examples for words that showed more variation in the cognate analysis.

<table>
<thead>
<tr>
<th>Language</th>
<th>Word</th>
<th>Word</th>
<th>Word</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brokskat</td>
<td>ʃæʒiri</td>
<td>χalma</td>
<td>ruŋ</td>
<td>myʃ</td>
</tr>
<tr>
<td>Kalkoti</td>
<td>dʒeːmel</td>
<td>dʒagar</td>
<td>kʰan</td>
<td>meʔʃ</td>
</tr>
<tr>
<td>Kohistani Shina</td>
<td>fœːː z̪ː</td>
<td>hjiːː</td>
<td>kʰoŋ</td>
<td>muːʃaː</td>
</tr>
<tr>
<td>Kundal Shahi</td>
<td>saːl</td>
<td>kledʑe</td>
<td>batoː</td>
<td>mɛːfɛː</td>
</tr>
<tr>
<td>Palula</td>
<td>wiːwij</td>
<td>dʒʰaŋːr</td>
<td>kʰaːŋ</td>
<td>mɪʃ</td>
</tr>
<tr>
<td>Sawi</td>
<td>aıyaj</td>
<td>jìːnuː</td>
<td>dʑaːʔu</td>
<td>manuːʃ</td>
</tr>
<tr>
<td>Gilgiti Shina</td>
<td>sala</td>
<td>dʒiːgar</td>
<td>tʃiːʂ</td>
<td>moʃaː</td>
</tr>
<tr>
<td>Gurezi Shina</td>
<td>saːroʔu</td>
<td>sina</td>
<td>pahaːʔ</td>
<td>nafar</td>
</tr>
<tr>
<td>Ushojo</td>
<td>ŋiːjoː</td>
<td>dʒagar</td>
<td>kʰoːn</td>
<td>manuːʒuː mufaː</td>
</tr>
</tbody>
</table>

5.2 Feature Analysis

5.2.1 Phonology

Syllable Structure

Five of the Shina languages, namely Gilgiti Shina, Gurezi Shina, Kalkoti, Sawi and Kohistani Shina have moderately complex syllable structures. The other four, Brokskat, Kundal Shahi, Palula and Ushojo have complex syllable structures. The results are summarized in Table 5.

Table 5: Results of the syllable structure analysis.

<table>
<thead>
<tr>
<th>Syllable structure</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>simple</td>
<td>-</td>
</tr>
<tr>
<td>moderately complex</td>
<td>Gilgiti Shina, Gurezi Shina, Kalkoti, Sawi, Kohistani Shina</td>
</tr>
<tr>
<td>complex</td>
<td>Brokskat, Kundal Shahi, Palula, Ushojo</td>
</tr>
</tbody>
</table>

For Kalkoti and Sawi, three types of syllables (CV, CVC and CCV) were found, as illustrated in the examples (3) and (4).

(3) Kalkoti syllable structures

a. CV, e.g. [duː] "two"

b. CVC, e.g. [mur] "die"

c. CCV, e.g. [traː] "three"

(4) Sawi syllable structures

a. CV, e.g. [suːri] "sun" (CV.CV)

b. CVC, e.g. [raːt] "blood"

c. CCV, e.g. [trojiʃ] "13" (CCV.CVC)

These results speak for a moderately complex syllable structure. Liljegren (2013: 136) found, in a brief analysis of Kalkoti, that word-initial two-consonant clusters seem to be restricted to /tr/ or /dr/, which confirms its classification as moderately complex. Buddrus (1967: 17) only discusses consonant clusters in his description of Sawi, pointing out that in the onset, two-consonant clusters with /j/ or
/r/ are allowed, and in the coda, two-consonant clusters with a nasal + consonant are allowed. These results confirm my own findings, and I would therefore argue that Sawi has a moderately complex syllable structure.

In the data for Kohistani Shina, Gilgiti Shina, and Gurezi Shina, only two types of syllables were found - CV and CVC. Examples that speak for a moderately complex syllable structure are given in (5) and (6).

(5) Kohistani Shina syllable structures
   a. CV, e.g. [suːr] “sun” (CV.CV)
   b. CVC, e.g. [leːl] “blood”

(6) Gilgiti Shina syllable structures
   a. CV, e.g. [ʂa] “full”
   b. CVC, e.g. [piːk] “drink”

(7) Gurezi Shina syllable structures
   a. CV, e.g. [kuʈu] “ear” (CV.CV)
   b. CVC, e.g. [leːl] “blood”

Schmidt et al. (2008: 28-29) do not discuss the syllable structure of Kohistani Shina in the way that it is discussed in this analysis (their focus lies on stress and pitch contrasts instead), but it was also not possible to find any other structures in their numerous examples, than the ones given in example (5) I would therefore argue that Kohistani Shina has a moderately complex syllable structure.

Radloff (1999: 45-49) also confirmed the existence of CV, CVV, CVC, CVVC and even CCV syllables in Gilgiti Shina. However, her data suggests that two-consonant clusters in the onset are limited to plosive + /r/, which means that Gilgiti Shina has a moderately complex syllable structure.

Ahmed (2016) does not discuss the syllable structure of Gurezi Shina, so it was not possible to compare my results to his. However, among the examples in his paper, the most elaborated syllable structure was CCVC in /braːn/ “to fall down” (2016: 101). Depending on whether such consonant clusters in the onset are restricted, this could speak for either a moderately complex or complex syllables structure. As only one example was found in the whole paper, and it does fit into the pattern for onset consonant clusters in Gilgiti Shina (plosive + /r/), it is likely that Gurezi Shina has the same type of limitations, and can therefore also be classified as having a moderately complex syllable structure.

The rest of the languages in the sample have complex syllable structures. For Brokskat, four types of syllables were found, as illustrated in example (8).

(8) Brokskat syllable structures
   a. CV, e.g. [aːʈi] “bone” (V.CV)
   b. CVC, e.g. [dud] “breast”
   c. CCV, e.g. [njæ] “fish”
   d. CCVC, e.g. [skis] “see”

Ramaswami (1982: 31-37) had similar results, but also found some occurrences of CCCV(C), such as [sbraːŋt͡si] “honey” (1982: 32). This syllable type did not occur very frequently - it might be a loan from Tibetan languages - however, it aligns with the results from the field data, which show that Brokskat has a complex syllable structure. Sharma (1998: 48-50) confirms these findings as well.

In the Kundal Shahi data, four types of syllables were found, as example (9) illustrates.
Kundal Shahi syllable structures
a. CV, e.g. [piː] “drink”
  b. CVC, e.g. [kaːn] “ear”
  c. CCV, e.g. [kleːd͡ʑe] “liver” (CCV.CV)
  d. CVCC, e.g. [daːnd] “tooth”

Baart and Rehman (2005: 11) briefly discuss possible consonant clusters, pointing out that they are most likely restricted to plosive + /r/ (initial and non-initial), and sibilant + plosive or nasal + plosive (final). These findings confirm that Kundal Shahi has a complex syllable structure, as two-consonant clusters are allowed in the coda.

For Palula, the syllable types that are illustrated in example (10) were found.

(10) Palula syllable structures
a. CV, e.g. [nòː] “name”
  b. CVC, e.g. [haɳɖúk] “bone” (CVC.CVC)
  c. CCV, e.g. [troː] “three”
  d. CCVC, e.g. [triːʃ] “thirteen”
  e. CVCC, e.g. [ʂiŋɡ] “horn”

These results speak for a complex syllable structure, as two-consonant clusters are allowed in the coda. Liljegren (2016: 64-69) confirms that CV, CVV and CV(V)C syllables are very common in Palula, and even onsetless syllables such as V(C) and VV(C) are possible. He also found clusters of two or three consonants in the onset, and two consonants in the coda, which confirms that Palula has a complex syllable structure.

For Ushojo, three syllable types, which are exemplified in (11) were found.

(11) Ushojo syllable structures
a. CV, e.g. [muːru] “die” (CV.CV)
  b. CVC, e.g. [kʰoːn] “mountain”
  c. CVCC, e.g. [doːnd] “tooth”

The two-consonant cluster in the coda of the last example speaks for a complex syllable structure. Decker (1992) does not discuss the syllable structure of Ushojo, but in her numerous examples more two-consonant clusters were found. Examples are given in (12).

(12) More examples for Ushojo syllable structures (Decker, 1992: 201-203)
  a. [mʌrg] “kill” (CVCC)
  b. [brʌbʌr] “connect” (CCV.CVC)
  c. [krom] “work” (CCVC)

Again, a two-consonant cluster occurs in the coda, which confirms that Ushojo has a complex syllable structure.

Aspiration

The analysis showed that most Shina languages still have a phonemic contrast for voiceless aspirates. The results are summarized in Table 6. However, in almost all languages, it was noticeable that this contrast is about to be replaced by e.g. rising tone, devoicing of the following vowel, clusters with /h/ or even a complete reinterpretation into a different phoneme (e.g. [pʰ] → [f]). In general, it was difficult to categorize the languages according to the binary values that were proposed in the beginning.
- this matter will be discussed further in 6.2. Only Palula has a phonemic contrast for the voiced aspirates, but some traces of it were still recognizable in some of the other languages as well. Again, in most Shina languages, the underlying contrasts have been replaced by e.g. tonal contrast, or have been reinterpreted into clusters with /h/.

<table>
<thead>
<tr>
<th>Features</th>
<th>Languages where the feature is present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless and voiced aspirates</td>
<td>Palula</td>
</tr>
<tr>
<td>Only voiceless, no voiced aspirates</td>
<td>Brokskat, Kalkoti, Kohistani Shina, Gilgiti Shina, Gurezi Shina, Kundal Shahi, Ushojo</td>
</tr>
<tr>
<td>No aspiration</td>
<td>Sawi</td>
</tr>
</tbody>
</table>

Table 6: Results of the aspiration contrast analysis.

In the Brokskat data, only the voiceless aspirated consonants [pʰ], [tʰ], [ʈʰ] and [kʰ] were found, as well as one minimal pair proving the phoneme status of [tʰ], which is given in example 13.

(13) [aːʈʰiː] 'bone' - [aːʈʰiː] 'eye'

Sharma (1998: 36-38) found six voiceless-aspirated consonants: [pʰ], [tʰ], [ʈʰ], [cʰ], [kʰ] and [tʃʰ]. Ramaswami (1982: 7-9) found the same set of aspirates, except for the palatal plosive /cʰ/. Neither of them discussed the topic of voiced aspirates. Ramaswami (1982) only gives a few examples of [驸], but could not prove that it was a phoneme on its own. Sharma (1998) did not find any occurrences of it. We can therefore state that Brokskat only has an aspiration contrast for voiceless consonants.

In the Kalkoti data, the voiceless-aspirated consonants [kʰ], [pʰ] and [tʰ] occurred, as well as traces of the voiced-aspirated consonants [bʰ] and [rʰ]. However, I would interpret these possible voiced-aspirated consonants as breathy voice in the following vowel, not aspiration. There appeared to be rising tone in these examples as well. This suggests that Kalkoti possibly had a phonemic contrast for voiceless aspirated consonants, as most IA languages, but has lost, or is in the process of losing, the contrast for voiceless consonants, as they were so few and difficult to notice. Instead, these contrasts might have been replaced by e.g. tonal contrast. This also agrees with the results of Liljegren (2013: 129) - according to him, Kalkoti still has a phonemic contrast for voiceless aspirates, but the original voiced aspirated consonants have disappeared (Liljegren, 2013: 136), and are now reflected by a low tone (Liljegren, 2013: 143).

In Kundal Shahi the aspiration was in general difficult to hear, not only for voiced consonants, but also for voiceless. Many examples, where aspirated consonants would usually have appeared, the languages seemed to use rising tone to mark the contrast instead. Baart and Rehman (2005: 10) only recognized a phonemic contrast for voiceless-aspirated consonants, arguing that it is likely that the contrast for voiced-aspirates has already been lost, as in many other IA languages in the Hindu Kush region. When comparing some of their examples, such as [kaːʈʰoː] 'wood' (Baart and Rehman, 2005: 10) to the examples in the field data, the aspiration was not as clear as the authors described it. This may simply be due to the fact that my analysis only relies on one speaker, which makes it difficult to explore the topic any further. These problems are discussed in more detail in 6.2. Considering my own findings, and the previous study by Baart and Rehman (2005), I would therefore argue that Kundal Shahi only has a weak aspiration contrast for voiceless-aspirated consonants left, which is in the process of being replaced by e.g. a tonal contrast, and for voiced-aspirated consonants, is has already been lost entirely.

For Palula, both voiceless-aspirated consonants ([pʰ], [kʰ]) and voiced-aspirated consonants ([dʰ]) were found. In addition, one minimal pair for the contrast between [k] and [kʰ] could be identified, as illustrated in example 14.
Therewerealsotracesofothervoiced-aspiratedconsonants,butinmostcasesthecontrastsappearto
havebeenreinterpretedwithrisingtone. One ofthefewclearexamplesofvoicedaspirationwas[dʰiː] 'daughter’. IwouldthereforearguethatPalulaclearlyhasacontrastforvoicelessandvoicedaspirates,
butthecomparesforthevoicedaspiratesshowacleartendencytobereinterpretedinotherways.

Liljegren(2016: 69-70)alsorecognisedaphonemiccontrastforvoicelessplosivesandaffricates,and
regardingthethotheadspireshearguesthathertisthisclearcontrast,especiallywhencomparing
tocognatesinothermajorShinavarieties,eventhoughhepreferredtoanalyzethemast voiceclustering
with /h/. The main reason for that is that in individuallexemes, /h/ is movable - varying from speaker
to speaker (Liljegren, 2016: 70-71). As an example, he gives the word for 'elder', which sometimess
pronounced [ghaɖeːro], and sometimess [gaɖheːro] (Liljegren, 2016: 71).

ForGilgitiShina,onlyafewoccurrencesof[pʰ]and[ʈʂʰ]werefound, butnominalpairsthrowing
theirphonemestatus. In some cases, [pʰ] was also pronounced [f]. Looking at cognates in other
languages, where voiceless-aspiratesoccur, there does not seem to be any aspiration left inGilgiti
voiceless-aspiratedconsonantsandealsodothestatusofthevoicedaspirates, statingthatitis
likelythatGilgitiShinahasalreadylostthem,consideringthefactthatisthisatypicalfeatureof
Dardiclanguages. Degener and Zia (2008: 14)reachedthesameconclusion. I wouldthereforeargue
thatGilgitiShinastillhasaphonemiccontrastforvoicelessaspirates, even thoughitisrecognizable
thatthiscontrastisabouttobereplacedbyothercontrasts, as e.g. [pʰ] being pronouncedit like [f] in-
stead. Thecontrastforvoice-aspiratedconsonantshasalreadybeenreplacedentirely,possiblybya
tonalcontrast.

ForGureziShina, the twovoiceless-aspiratedconsonants[pʰ] and[ʈʂʰ]werefound, butagainwithnom
minimalpairs. Mostly, the aspiration was difficult noticende, and in some cases I even heard a tendency
topronounce the [pʰ]-sound rather like [f]. Ahmed (2016: 95)provedthsexistenceofsevenvoiceless-
aspiratedconsonantphonemeswithminimalpairs. Thus, I would suggest that Gurezi Shina still has
aphonemiccontrastbetweenvoicelessandvoiceless-aspiratedconsonants, even though it might be
abouttobereplacedbyothercontrastsine.g. thebilabialposition.

IntheUshojodata, the twovoiceless-aspiratedconsonants[pʰ] and[kʰ]occurred, as well astwom
minimalpairsprouvingtheircontrastwith [p] and [k] respectively. Theexamplessagivenin 15 and

(15) [kʰoːn] ‘mountain’ - [koːn] ‘ear’

(16) [pʰo] ‘son’ - [poːn] ‘path’

Decker (1992: 73)alsofoundasetofvoicelessaspiratedconsonants. Shaelso discussed thestatus
ofvoicedaspirates, reachingtheconclusionthattheycannotbetreatedas singlephonemesanymore,
especiallysince manyDardiclanguageshavelostthem. IthereforeconcludethatUshojohasaphonemic
contrastforvoiceless-aspiratedconsonantsleft, butnotanymoreforyciced-aspiratedconsonants.

ForKohistaniShina, onlyvoiceless-aspiratedconsonants were found ([pʰ], [tʰ], [ʈʰ], [ʈʂʰ] and[kʰ]),
aswellasaminimalpairfor[k] - [kʰ] that is given in 17.

(17) [kʰoːn] “mountain” - [koːn] “ear”

Asin some otherShinalanguages, there was clearlyatendencytoreplacementheaspirationcontrastwith
adevoicingofthefollowingvowel. Schmidt et al. (2008: 18-21)foundvoiceless- and voiced-aspirated
consonants, butalsonoticedthatvoiced-aspiratesonlyoccurin in thebeginningofwords, andonly
in loanwords. They argued that the Old Indo-Aryan (OIA) voiced aspirates have been lost, as inmany
otherHKIALanguages, and thatvoiced-aspirates are arathernewdevelopmentthatcameintothe
language via loan words. This also explains why no voiced aspirates occurred in the data - the questionnaires were specifically designed with the intention to avoid words that could be loans from other languages, using words that are typically considered being basic vocabulary. We can therefore state that Kohistani Shina does have an aspiration contrast for both voiceless- and voiced consonants, but the latter do only occur in loan words, and not in the "original" vocabulary. As I am trying to rely my analysis mainly on basic vocabulary, and not loan words, the value for Kohistani Shina will still be no for the voiced aspirates.

In the Sawi data, only the aspirated consonant [pʰ] was transcribed, which was also very difficult to judge whether it could even be counted as aspirated. It could also be perceived as a re-interpretation into [f]. The cognate in Gawarbati was not aspirated either, and in general there were no traces of aspiration in the Gawarbati data either. However, Buddruss (1967: 15) found both voiceless- and voiced-aspirated consonants - stating that he was rather sure about the phoneme status of the voiceless-aspirated consonants, especially the plosives, and less sure about the voiced aspirates. Considering that Buddruss' description is more than 50 years older than the data for this study, I would argue that Sawi has lost its aspiration contrasts entirely during that time (or rather re-interpreted it into [f], the contrast with [p] still exists), and he was already able to witness how they started to disappear in the voiced consonants. In this regard it has of course to be considered that this change only happened at one place of articulation (bilabial), without affecting the others, as e.g. in Gurezi Shina. However, in the case of Sawi that does not seem likely, especially when comparing other words to cognates in other Shina languages, that still have aspiration.

**Retroflexion**

The analysis showed that most Shina languages have retroflex plosives and fricatives, as most HKIA languages (see Table 7). The only exception is Kundal Shahi, which is missing the retroflex fricatives. The data for the western Shina languages Palula, Sawi and Kalkoti suggests that they do not have the voiced-retroflex fricative [ʐ] - or that its phoneme status is unclear.

<table>
<thead>
<tr>
<th>Retr. plosives</th>
<th>Retr. fricatives</th>
<th>Other retr. consonants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>[ʂ]: Palula, Sawi, Kalkoti</td>
<td>[ʈ]: Brokskat</td>
</tr>
<tr>
<td></td>
<td>[ʂ], [ʐ]: Brokskat, Gilgiti Shina, Gurezi Shina, Ushojo, Kohistani Shina</td>
<td>[ʈ, ɳ]: Kundal Shahi</td>
</tr>
<tr>
<td></td>
<td>Kohistani Shina</td>
<td>[ʈ, ɳ, ʂ]: Kohistani Shina, Palula, Sawi, Gilgiti Shina, Gurezi Shina, Ushojo</td>
</tr>
<tr>
<td>Absent</td>
<td>none</td>
<td>Kundal Shahi</td>
</tr>
<tr>
<td></td>
<td>Kalkoti(?)</td>
<td>Kalkoti(?)</td>
</tr>
</tbody>
</table>

In the Brokskat data, two retroflex plosives ([ʈ], [ʈʰ]) and two retroflex fricatives ([ʂ], [ʐ]) were found, as well as one minimal pair for the contrast between [s] and [ʂ], which is given in (18).

(18) [sat] “seven” - [ʂa] = “six”

Furthermore, the retroflex tap [ʈ] occurred in a few examples. Ramaswami (1982: 7) named the same retroflex fricatives that occurred in the current analysis, as well as a set of three alveolar plosives, that contrast with the dental [ʈ], [ʈʰ] and [ɖ], where I would assume that these are just another interpretation of the alveolar - retroflex contrast that I heard. The exact place of articulation of alveolar, post-alveolar and retroflex consonants can sometimes be difficult to determine, but the important part is that a contrast was recognized. Sharma (1993: 36-37) recognized the same set of retroflex plosives ([ʈ], [ʈʰ], [ɖ]) as Ramaswami (1982), as well as the two retroflex fricatives found in the data. I would therefore argue that Brokskat has both retroflex plosives and fricatives, as well as the retroflex tap [ʈ]. In one example, [ɳ]
did also occur, but as it was the only one, and Sharma (1998: 36) was also not sure whether it actually has the status of a phoneme, I decided not to count it into the inventory.

For Kohistani Shina, the retroflex plosives [ʈ] and [ʈʰ], as well as the retroflex fricatives [ʂ] and [ʐ] were found. A near-minimal pair illustrating the contrast between [s] and [ʂ] was found as well, as illustrated in 19.


[ʈ], [ɳ] and [ʈʂ] also occurred in the data. Schmidt et al. (2008: 16), also recognized retroflex plosives (ʈ/ʈʰ/, /ɖ/) and fricatives (/ʂ/, /ʐ/), confirming that Kohistani Shina has both types of retroflex consonants. They also confirm the existence of /ʈ/, /ɳ/ and /ʈʂ/ as phonemes in Kohistani Shina (Schmidt et al., 2008: 18-23).

For Gilgiti Shina, the following retroflex plosives and fricatives were identified: [ʈ], [ɖ], [ʂ], [ʂʰ] and [ʐ]. Furthermore, a minimal pair for the contrast between [s] and [ʂ] was found, as illustrated in Example 20.

20) [sa] “six” - [sa] “sister”

Regarding the existence of any other retroflex consonants, [ʈ], [ɳ] and [ʈʂ] occurred in the data, and were also confirmed by Radloff (1999: 22). Furthermore, Radloff (1999: 22) identified three retroflex plosives (ʈ/ʈʰ/, /ɖ/) and two retroflex fricatives (/ʂ/, /ʐ/), and she confirmed their phonemic status with minimal pairs (1999: 24, 28). Thus, we can say that Gilgiti Shina has retroflex plosives and fricatives, as well as [ʈ], [ɳ] and [ʈʂ].

In the Gurezi Shina data, only two retroflex plosives ([ʈ], [ʈʰ]) and one retroflex fricative ([ʂ]) were found. In addition, [ʈ], [ɳ] and [ʈʂ] occurred. Example 21 illustrates the contrast between [s] and [ʂ] and in word-initial positions.

21) [sa] “six” - [sas] “sister”

Ahmed (2016: 95-96) found both retroflex plosives (ʈ/ʈʰ/, /ɖ/) and fricatives (/ʂ/, /ʐ/), which is why I would argue that Gurezi Shina has those two types of retroflex consonants. They also confirmed the same set of additional retroflex plosives ([ʈ], [ɳ] and [ʈʂ]) as phonemes.

In the Ushojo data, the following retroflex plosives and fricatives were found: [ʈ], [ʈʰ], [ɖ], [ʂ], [ʐ], [ʐʰ]. Furthermore, two examples for the contrast between [s] and [ʂ], as well as [z] and [ʐ] were found (see example 22 and 23).

22) [sas] “sister” - [ʂa] “six”

23) [zar] “one thousand” - [za] “brother”

Decker (1992: 73) also recognized retroflex plosives (ʈ/ʈʰ/, /ɖ/) and fricatives (/ʂ/, /ʐ/), thus I would argue that Ushojo has both retroflex plosives and fricatives. As in most other Shina languages, [ʈ], [ɳ] and [ʈʂ] also occurred in Ushojo, which was also confirmed by Decker (1992: 73).

In the Palula data, four retroflex consonants were found: [ʈ], [ɖ], [ʂ] and [ʐ]), but there was not enough data to find minimal pairs for any of them. Liljegren (2016: 53-54), however, does recognize /ʈ/, /ʈʰ/, /ɖ/ and /ʂ/ as phonemes. He also mentioned the existence of [ʐ] (which did not occur very frequently), and was not sure whether there was a phonemic contrast with the voiced retroflex affricate. According to him, it is more likely that [ʐ] and [ɖʐ] are environmentally conditioned allophones of the same phoneme (Liljegren, 2016: 54). Nonetheless, I would argue that Palula has both retroflex plosives and fricatives, even though the phonemic status of the voiced-retroflex fricative [ʐ] is uncertain. Also, the additional retroflex consonants [ʈ], [ɳ] and [ʈʂ] occurred in the data as well, and were also confirmed in Liljegren’s grammar (Liljegren, 2016: 53).

For Sawi, the retroflex consonants [ʈ], [ɖ] and [ʂ] were found, as well as a near-minimal pair illustrating the contrast between [d] and [ɖ] in word-initial positions (see example 24).
(24) [ɖaɖu] “mountain” - [da:ndu] “tooth”

Buddrus (1967: 15) recognized the retroflex plosives /ʈ/, /ʈʰ/ and /ɖ/, as well as the retroflex fricative /ʂ/. He also pointed out the existence of [ʐ], but was not sure about its phonemic status. He identified the retroflex fricatives as post-alveolar consonants, but as there is still a clear contrast to the alveolar and palatal phonemes, I would again argue that this is just a question of different interpretation of the same sounds - the difference between post-alveolar and retroflex can sometimes be difficult to hear. I would therefore say that Sawi has retroflex plosives and fricatives, even though the phonemic status of [ʐ] is unclear. As in most other Shina languages, [ɽ], [ɳ] and [ʈ͡ʂ] occurred as well, and were also confirmed by Buddrus (1967: 15).

For Kalkoti, only two retroflex consonants were found: [ʈ] and [ʂ]. Example (25) indicates that there is a contrast between [s] and [ʂ].

(25) [siːr] “sun” - [ʂiŋ] “horn”

Liljegren (2013: 135) found, in a preliminary analysis of Kalkoti, the retroflex plosives /ʈ/, /ʈʰ/, and /ɖ/, as well as the voiceless-retroflex fricative /ʂ/. The fact that [ʐ] seems to be missing, as opposed to most other Shina languages, agrees with Palula and Sawi, where the phonemic status of it also is unclear. Accordingly, I would argue that Kalkoti has retroflex plosives, and a voiceless-retroflex fricative. Regarding the other retroflex consonants, there was quite a lot of uncertainty. [ʈ] did occur quite frequently in different contexts, but according to Liljegren (2013: 135), its phoneme status is not completely clear, considering that its pronunciation varies a lot, making it hard to find the underlying phonemes, and that it might have come into the language through loans from Pashto and Urdu. Regarding the phoneme status of [ʈ͡ʂ], Liljegren (2013: 135) reached a similar conclusion as for [ʈ]. [ɳ] did only occur in combination with [ɖ], which is why I would assume that this is just a case of place assimilation, and that [ɳ] is not a phoneme on its own. Liljegren (2013: 135) also reasons along the same lines.

For Kundal Shahi, only retroflex plosives were found: [ʈ], [ʈʰ] and [ɖ], but no fricatives. Baart and Rehman (2005: 10) found exactly the same set of retroflex plosives, and no retroflex fricatives either. The latter are in fact typical for “Dardic” languages, but the authors point out that for example Kashmiri, another IA language spoken in the area, does not have them either. We can therefore conclude that Kundal Shahi does only have retroflex plosives, but not retroflex fricatives. [ʈ] and [ɳ] also occurred in the data, and were also recognized as phonemes by Baart and Rehman (2005: 10).

5.2.2 Morphology

Grammatical Gender

All Shina languages in the sample have grammatical gender, and all of these systems are sex-based. None of them showed any indications of an additional animacy-based gender system. The results are presented in Table 8.

<table>
<thead>
<tr>
<th>Grammatical gender</th>
<th>Sex-based gender</th>
<th>Animacy-based gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>present</td>
<td>all</td>
<td>none</td>
</tr>
<tr>
<td>absent</td>
<td>none</td>
<td>all</td>
</tr>
</tbody>
</table>

For Brokskat, Kohistani Shina, Kundal Shahi, Palula, Sawi, Gilgit Shina, and Gurezi Shina, sufficient examples of adjectives, that agree with the noun in gender were found in the valency questionnaire, as illustrated for Kohistani Shina and Palula in example 26 and 27. All of the gender systems are based on biological sex.
The results align with findings from previous studies: Sharma (1998: 56) found that animate nouns in Brokskat have a masculine-feminine gender distinction, which is visible on the noun itself, as well as on pronouns, adjectives, and verbs. According to Schmidt et al. (2008: 40), Kohistani Shina nouns have masculine or feminine gender, and adjectives with vowel-endings agree with it (2008: 100), as well as verbs (2008: 58). Kundal Shahi, Sawi, and Gurezi Shina also have sex-based gender systems where verbs and adjectives agree with the nouns (Baart and Rehman 2005: 13; Buddrus 1967: 29; Qadir and Anand 2018: 15). According to Liljegren (2016: 84), Palula has a sex-based gender system, where e.g. adjectives and present tense verbs agree with the noun. For Gilgiti Shina, Degener and Zia (2008: 18-19) found similar results: adjectives and verbs that agree with nouns in gender.

For Ushojo, grammatical gender has never been analyzed in previous studies, so here I had to rely entirely on the field data. The examples showed a sex-based grammatical gender, as in the other Shina languages. Example 28 illustrates the distinctions.

(28) Ushojo, (v-47, v-77)

a. dq̴ag-u manuzu tuχi̴l-u
   old-m man.m cough.pst(-)-m
   'The old man coughed.'

b. dq̴ag-i dʒapaji p̴ija: ti gaŋ suŋa: t̴e-g-i
   old-f woman.f girl to song hearing do-pfv-f
   'The old lady taught the girl a song.'

In Kalkoti, grammatical gender is not marked with a suffix on the adjective, as in the other Shina languages, but instead a vowel in the stem is modified, according to whether the noun is masculine or feminine. The examples 29 and 30 illustrate the gender agreement.

(29) Kalkoti, (v-47, v-77)

a. ganar me̴l̴ k̴usil
   old.m man.m cough.pst
   'The old man coughed.'

b. buq̴-j̴ kuŋi p̴aj t̴e r̴oŋ izda t̴awoŋ-o
   old-f woman.f girl to song.m learning make.do pfv-m.sg
   'The old lady taught the girl a song.'
b. *gan:*r pe: *tah* ro: bʰudʒaːl
   old woman.f girl.f to song?

'The old lady taught the girl a song.'

(30) Kalkoti, (k-10, k-12)

a. *log drah*
   young.m brother.m

'younger brother'

b. *leg ban*
   young.f sister.f

'younger sister'

**Morphosyntactic Alignment (full noun phrases)**

According to Liljegren (2017: 132-133), all languages in the sample, except for Sawi and Gurezi Shina, have ergative-absolutive alignment for noun phrases. Sawi has a tripartite alignment. As Gurezi Shina was not included in Liljegren's sample, other data sources had to be considered. Ahmed (2019: 89) found a tripartite case alignment in Gurezi Shina as well. The results are summarized in Table 9.

<table>
<thead>
<tr>
<th>Alignment type</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accusative-nominative</td>
<td>none</td>
</tr>
<tr>
<td>Ergative-absolutive</td>
<td>Brokskat, Palula, Kalkoti, Gilgit Shina, Kohistani Shina, Kundal Shahi, Ushojo</td>
</tr>
<tr>
<td>Tripartite</td>
<td>Sawi, Gurezi Shina</td>
</tr>
</tbody>
</table>

For the languages with an ergative-absolutive alignment, a variety of examples were found in the field data, as illustrated in 31a and 32a for Gilgit Shina. Example 31a shows that A, the subject of a transitive clause, is marked with ergative case, whereas S, the subject of an intransitive clause (31b), is not marked. Example 32 shows that the object of a transitive clause (O), and the subject of an intransitive clause (S) are not marked.

(31) Gilgit Shina (v-85, v-70)

a. *muʃa*-s gafijak sanagow
   man-erg stick make.pst
   A   O

'The man made a stick.'

b. *ano muʃa: daruːʃ han*
   This man be.m.sg hunter
   S

'This man is a hunter.'

Important to note is here that the examples in 31b and 32b actually not are intransitive, but rather examples of non-verbal predication. However, they still confirm the alignment patterns.
a. maː-s  baːl  dudʒareːgin
   mother-erg  baby  wash.pst
   A  O
   'The mother washed the baby.'

b. baːl  ojano  han
   baby  hungry  be.m.sg
   S
   'The baby is hungry.'

Regarding case alignment in Gurezi Shina, Ahmed (2019: 89) found that the language has a tripartite case alignment, where S is not marked (which equals nominative case in this language), O is marked with oblique case, and A is marked with agentive case (which has different forms for perfective and imperfective forms). However, he also noticed a tendency to leave out the marker for oblique case, assuming that this case is about to disappear (Ahmed, 2019: 33). This tendency is also visible in the field data, as example 33 illustrates: the oblique case of [muːlaj] ‘girl’ in b is not marked.

(33) Gurezi Shina (v-57, v-3, v-37)
   a. t͡ʃuːɳiː  muːlaj-Ø  haziː-l-iː
      little.f  girl-nom  laugh-pfv-3.f.sg
      S
      'The little girl laughed.'
   b. baːl-iː  muːlaj-Ø  waːriː  ṭ͡ʂakow
      boy-ag.pfv  girl-obl  toward  see.pst.3.m.sg
      A  O
      'The boy looked at the girl.'
   c. muːlaj-ow  toːmiː  daːjd-eːɖ  pʰuːʃ-i  ṭ͡ʂiːɳ-iː
      girl-ag.pfv  refl  grandmother-?  flower-pl(?)  send.f
      A  O
      'The girl sent flowers to her grandmother.'

Sawi has a tripartite case alignment, according to Liljegren (2017: 132-133). Unfortunately, it was difficult to find sufficient examples in the field data and gloss them correctly.

5.2.3 Syntax

Word Order

All Shina languages in the sample have a SOV word order, as the examples 34 and 35 for Brokskat and Kalkoti show.

(34) Brokskat (v-01)
    bjøː  kʰazas  χat
    boy  fruit  eat.pfv
    'The boy ate the fruit.'
Adpositions

All languages in the sample have postpositions which follow the noun phrase (NP). Sometimes it was difficult to determine whether the languages have postpositions or case-suffixes, and due to the limited data it was not possible to conduct a detailed analysis. However, for some languages, this question has been discussed previously, otherwise it was necessary to rely on whether the words seemed to be phonologically distinct or not. The data for Gilgiti Shina and Gurezi Shina suggested that these languages preferred locative case suffixes over postpositions, but in previous studies, postpositions have been found in these languages as well (*Degener and Zia*, 2008: 15-16, *Ahmed*, 2019: 45).

For Brokskat, only few examples of postpositions were found, as there seems to be a tendency to use case suffixes instead. Still, a few postpositions were identified, an examples is given in (36).

(36) Brokskat, (v-84)

\[
\text{pala tʃaqmaʃ noː naχazoː ut}
\]

\[
\text{apple tree from? fall ?}
\]

'The apple fell from the tree.'

For Ushojo, there was also no information about adpositions in previous studies, so here the only data source was the field data. Again, a few examples of postpositions were found, one of them is given in (37).

(37) Ushojo (v-42)

\[
\text{manuʐuː gilaːs dar wuj uliːgi}
\]

\[
\text{man glas into water pour.rst}
\]

'The man poured water into the glass.'

For Kundal Shahi, Palula, Sawi, Kalkoti and Gurezi Shina, evidence for postpositions was found both in the data and in previous descriptions. *Baart and Rehman* (2005: 14) found that Kundal Shahi has postpositions which tend to follow the oblique form of the noun. They gave examples such as almaaríi laa ‘next to the cupboard’, and kǘr̻ sam ‘with the woman’. Example (38) shows the instrumental postposition sam.

(38) Kundal Shahi (v-27)

\[
\text{kuɽun ɖiŋgiːr sam saːʁriːs daguːn}
\]

\[
\text{boy stick with snake beat.rst}
\]

'The boy beat the snake with a stick.'

*Liljegren* (2016: 92) found that Palula has postpositions such as the 'to, for', sangí 'with, at', and mají 'among, in, inside', which usually follow the noun in the oblique case. An example is given in (39).

(39) Palula (v-36)

\[
\text{asim kuŋːakoːm t'e kitiːbi dit-im}
\]

\[
\text{we children.obl.pl to book.f.pl gave.f.pl}
\]

'We gave the books to the children.'
Sawi has, according to Buddruss (1967: 45), many different postpositions, which stand with nouns in different cases. In the field data, one example was found, which is illustrated in (40).

(40) Sawi (v-48)

\begin{center}
\text{man\textsubscript{pl}} \text{ tree} \text{ up} \text{ climb}
\end{center}

The men climbed (up) the tree.'

According to Liljegren (2013: 147), Kalkoti has postpositions. In (41) an example from the field data is illustrated.

(41) Kalkoti (v-81)

\begin{center}
\text{boy} \text{ field} \text{ on} \text{ appear.pst}
\end{center}

'The boy appeared on the field.'

Gurezi Shina has, according to Ahmed (2019: 45), clitic and free postpositions. Example (42) shows the usage of the clitic postposition for 'in/at/on', which in this case could even be interpreted as locative case marker. The distinction between clitic postposition and locative case marker is difficult to make. Ahmed (2019) bases his analysis on the fact that the clitic postpositions do not have a pitch accent on their own, hence they are phonologically dependent on the head noun, which could also be an argument for analyzing them as locative case markers.

(42) Gurezi Shina (v-81)

\begin{center}
\text{woman} \text{ market-\text{cp}/\text{loc} go.pst}
\end{center}

'The woman went to the market.'

According to Schmidt et al. (2008: 70), Kohistani Shina has postpositions that usually stand with the oblique form of the noun, and in some cases with the nominative. An example for 'in' (43) was found in the field data:

(43) Kohistani Shina (v-42)

\begin{center}
\text{man} \text{ glas} \text{ water} \text{ pour(?)}
\end{center}

'The man poured water into the glass.'

For Gilgiti Shina, the data was rather difficult to interpret, as there were no phonologically distinct words which could be adpositions. Degener and Zia (2008: 15-16) discusses postpositions, case and derivational suffixes in Gilgiti Shina, and comes to the conclusion that only a few of these “formatives” can actually be seen as postpositions, the rest are case-suffixes. This would lead me to the conclusion that, even though Gilgit Shina does not have many freestanding adpositions, those few examples are postpositions. In the field data, most examples seemed to have case suffixes, as illustrated in (44).

(44) Gilgiti Shina (v-51)

\begin{center}
\text{children} \text{ bench-\text{loc} sit.pst}
\end{center}

'The children sat down on the bench.'
5.2.4 Lexicon

Numeral Bases and Numeral Composition

Table 10: Results of analysis of numeral bases.

<table>
<thead>
<tr>
<th>Numeral base</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Brokskat, Palula, Kalkoti, Gilgiti Shina,</td>
</tr>
<tr>
<td></td>
<td>Gurezi Shina, Kohistani Shina, Sawi,</td>
</tr>
<tr>
<td></td>
<td>Ushojo</td>
</tr>
<tr>
<td>10</td>
<td>Kundal Shahi</td>
</tr>
</tbody>
</table>

All languages in the sample, except for Kundal Shahi, have the numeral base 20. In these languages, the word for '30' is, for instance, composed with the words for '20' and '10'. '31' consists of '20' + '11', and so on (see Example 45).

(45) Palula numerals
   a. [daːʃ] '10'
   b. [akoːʃ] '11'
   c. [bʰiːʃ] '20'
   d. [bʰiːʃeː daːʃ] '30'
   e. [bʰiːʃeː akoːʃ] '31'

Kundal Shahi has the numeral base 10. Here, as in English, every tenth number has a new lexeme, as illustrated in example 46.

(46) Kundal Shahi numerals
   a. [dʰaj] '10'
   b. [biː] '20'
   c. [tɕiːɕ] '30'
   d. [tɕopeːr] '40'

Regarding the numeral composition, the Shina languages all showed the same pattern for constructions with 10, which was $x + 10$. Regarding the composition of constructions with 20, the languages showed more variation: 3 languages use $x + 20$ constructions, and 5 of them use $20 + x$. For Kundal Shahi, only constructions with 10 could be analyzed, as large parts were missing from the data. The results are summarized in Table 11.

Table 11: Results of analysis of numeral composition.

<table>
<thead>
<tr>
<th>Numeral composition</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x+10$ &amp; $x+20$</td>
<td>Kalkoti, Sawi, Ushojo</td>
</tr>
<tr>
<td>$x+10$ &amp; $20+x$</td>
<td>Brokskat, Kohistani Shina, Palula, Gilgiti Shina, Gurezi Shina</td>
</tr>
</tbody>
</table>

Example 47 illustrates the system for Gurezi Shina. Here, the number 11 is composed of [ak], which is derived from [ek], the word for '1', and [aːj], which is derived from [daj], the word for '10', resulting

---

*Kundal Shahi is missing in the table, as only constructions with 10 could be analyzed.*
in an x+10 construction. For the numbers from 20 onwards, the system flips: the component for '20' [bi:] comes first, followed by [ek] '1'.

(47) Gurezi Shina numerals
   a. [ek] '1'
   b. [daj] '10'
   c. [akaːj] '11'
   d. [biː] '20'
   e. [bigi ek] '21'

Example 48 illustrates the other system, x+10 and x+20, for Sawi. Here, the number 27 is composed of [saːt] '7' and [bij] '20', resulting in a x+20 pattern.

(48) Sawi numerals
   a. [saːt] '7'
   b. [deʃ] '10'
   c. [satijɑːʃ] '17'
   d. [biʃ] '20'
   e. [saːtɑːn bij] '27'

Kinship terms

The results of the analysis of kinship terms are presented in Table 12.

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Kinship Line</th>
<th>Direction of Kinship</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>present</td>
<td>Brokskat, Gilgiti Shina</td>
<td>Kalkoti, Kundal Shahi, Palula, Ushojo</td>
<td>Palula, Gilgiti Shina</td>
</tr>
<tr>
<td>absent</td>
<td>Kalkoti, Kohistani Shina, Kundal Shahi, Palula, Sawi, Gurezi Shina, Ushojo</td>
<td>Brokskat, Kohistani Shina</td>
<td>Shina, Kundal Shahi, Sawi, Gurezi Shina, Ushojo</td>
</tr>
</tbody>
</table>

Only two of the languages, Brokskat and Gilgiti Shina, lexicalize kinship line, meaning that they have a lexeme for 'parent', instead of saying 'mother and father'. This distinction is illustrated in example 49. Example 52, on the other hand, illustrates how the rest of the languages express the concept of 'parents': they simply say 'mother and father', instead of using a separate lexeme.

(49) Brokskat
   a. [boː] 'father'
   b. [aj] 'mother'
   c. [pʰæmæ] 'parent'

(50) Kalkoti
   a. [bʰab] 'father'
   b. [jeh] 'mother'
   c. [jeh teː bʰab] 'mother and father'
Regarding the direction of kinship, Brokskat, Kohistani Shina, Sawi, Gilgiti Shina and Gurezi Shina do not make a difference for e.g. 'father’s father’ and ‘mother’s father’ (see Example 51), whereas Kalkoti, Kundal Shahi, Ushojo and Palula do have different lexemes for that, as illustrated in example 52.

(51) Gilgiti Shina
   a. [duːdo] ‘father’s father’
   b. [duːdo] ‘mother’s father’

(52) Palula
   a. [dôːdo] ‘father’s father’
   b. [môːmo] ‘mother’s father’

Regarding the age contrast, only Palula and Gilgiti Shina do have different lexemes for e.g. ‘older brother’ and ‘younger brother’. The other Shina languages use combinations of words for e.g. ‘old’/ ‘big’ and ‘brother’. Examples are given in 53 and 54.

(53) Kohistani Shina
   a. [ʐâː] ‘brother’
   b. [báɽu ̥ ʐâː] ‘older brother’
   c. [lékʰu ̥ ʐâː] ‘younger brother’

(54) Palula
   a. [bʰrôː] ‘brother’
   b. [kâːku] ‘older brother’
   c. [kûːt͡ʃo] ‘younger brother’

Demonstratives

All languages in the sample have some kind of distance contrast for demonstratives. Five of the languages showed a 2-way distance contrast, and the other four have a distance contrast with 3 or more distinctions. The results are summarized in Table 13. All informants used sets of adnominal demonstratives, meaning that pronominal demonstratives were not considered in the analysis.

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-way</td>
<td>Kalkoti, Gilgiti Shina, Gurezi Shina, Sawi, Ushojo</td>
</tr>
<tr>
<td>3-way or more</td>
<td>Brokskat, Kohistani Shina, Palula, Kundal Shahi</td>
</tr>
</tbody>
</table>

Sawi, Gilgiti Shina, Gurezi Shina, Kalkoti and Ushojo showed a 2-way contrast. For Sawi the word [laː] was used to refer to the objects that were close by, whereas the ones that were further away were pointed out using constructions with [seː], [ta], or [tasiː], as illustrated in examples 55, 56.

Also possible 3-way contrast, but not confirmed in field data or previous studies.
(55) Sawi (d-6)

miː laː kitaːb χwax hinu:
1.SG PROX book is

'I like ___ book.'

(56) Sawi (d-10)

seː goʃ fubaːnuː hinu:
REM house beautiful M.SG is M.SG

'___ house is beautiful!'

Buddrus (1967: 40-43) translates [laː] with the German word 'dieser', which is used to denote an object that is nearby, and [seː] is translated with 'jener', denoting an object that is further away. These demonstratives are inflected for case and number, and they can be used as adjectives or nouns. Accordingly, [ta] and [tasiː] are accusative forms of the demonstrative for remote objects. Neither in the field data, nor in Buddrus (1967) grammar, there are any indications of a visibility contrast in Sawi demonstratives.

For Kalkoti, the demonstrative [ũː], for proximate objects, and [aːr] for remote objects were found in the field data. Examples are given in 57 and 58. There were no indications of a visibility contrast. Liljegren (2013:146) argues that there is a clear distance contrast between un and āār, which are adnominal-only demonstratives. Regarding a visibility contrast, he argues that it is likely that Kalkoti has such a contrast, as Palula has it as well, but this is something that has to be investigated in future studies. However, as there were no traces of such a contrast in the field data, Kalkoti will be classified as having a 2-way contrast.

(57) Kalkoti (d-5)

ũː kitaːb tʰi aː
PROX book your Q

'Is ___ your book?'

(58) Kalkoti (d-10)

aːr fiː baːt siliχ hin:
REM house much beautiful be.PRS

'___ house is beautiful!'

The informant for Gilgiti Shina used [ana], [ani], and [anu] when referring to closer objects, and the [pàreː] for remote objects, which indicates that there is a distance contrast. It is possible to assume that there also is a visibility contrast, as [pàreː] was only used when referring to the peak up in the hills and the village behind the hills, but there is not enough proof for that. Degener and Zia (2008: 22) also found a 2-way distance contrast: anú and ané were used when referring to something nearby, whereas ro and re were used to refer to something further away, as well as something not present and not visible. I would also argue that [pàreː] could be an older, longer, form of [re], but that they essentially have the same meaning. These results indicate that Gilgiti Shina most likely has a distance- and visibility contrast, even though the borders are still slightly unclear.

The results for Gurezi Shina were rather difficult to interpret, which was partially due to the fact that some sentences in the end were missing. However, the informant used the following words: [aniː], [anu], [ʐeː], and [aː]. According to Ahmed (2019: 38-39), Gurezi Shina has a 2-way contrast for demonstratives. anu is the pronoun used for proximate objects, and ʐo for remote objects. The author also points out that anu is often shortened to nu. As these pronouns are inflected for case and number,
other forms can occur: [aniː] is for instance the nominative plural form of anu. [aː] is not a demonstrative pronoun - it is a third person personal pronoun used for proximate objects/persons (Ahmed, 2019: 37-38). Those findings indicate that Gurezi Shina has a 2-way contrast for demonstratives.

The Ushojo informant mainly used [ana] for most of the examples, and [pila] for some of the objects that were further away, which indicates that there is a 2-way distance contrast. Decker (1992) also translated ana and ani with ‘this’, and pila and la with ‘that’, which indicates that she also found a distance contrast. There are no signs of a visibility contrast.

Brokskat, Kohistani Shina, Palula and Kundal Shahi have a 3-way (or more) contrast for demonstratives. For Brokskat, proximate and remote demonstratives were identified. For all closer objects, such as the ball hitting the finger, or the book lying on the table in front of them, the informant used the words [olo], [olom], [ala] and [alam] (see examples 59 and 60).

(59) Brokskat (n-1)
olom bolo mjo gülir tʰiʒid
prox ball my finger hit/touch?
’___ ball hit me on the finger.’

(60) Brokskat (n-8)
olpʰo polo tjo bita
rem ball your is/q?
‘Is ___ your ball?’

For everything that was further away, the informant used [olpʰo] instead. Furthermore, [ho] and [pari], which Sharma (1998) translated with ‘that’ and ‘there’ respectively, were used when referring to the objects that were the furthest away. Sharma (1998: 72) identified a set of adnominal demonstratives, which agree with the noun in gender and number. A summary is given in 61.

(61) Brokskat demonstratives, according to Sharma (1998: 72).

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>prox</td>
<td>homo</td>
<td>hamo</td>
<td>heme</td>
</tr>
<tr>
<td>rem</td>
<td>pʰo</td>
<td>pʰa</td>
<td>pʰe</td>
</tr>
</tbody>
</table>

To conclude, it is difficult to determine the exact meaning and usage of the demonstratives that occurred in the field data, especially as they differ from what has been named in previous studies, but it is clear that more than two demonstratives are used, depending on how far away the object is.

In the Kohistani Shina data, [aːe] and [aːj] were used when referring to the proximate objects, whereas [as] and [parəː] were used for the remote objects. Kohistani and Schmidt (2006: 88) found that Kohistani Shina has two demonstrative pronouns: aːe is used for proximate objects, and aːsə for remote objects. Both forms inflect for number and case. They do not name anything about a visibility contrast. [parəː] could be a third form, that is used for objects that are out of sight - it also seems to have cognates in other Shina languages, such as [parəː] in Gilgiti Shina, or [pari] in Brokskat. I would therefore argue that Kohistani Shina demonstratives with a 3-way distance and visibility contrast.

In the Palula data, four demonstratives, which agree with the gender of the noun, were found: [ani] and [anu], for proximate objects, as well as [aro] and [ape] for remote objects. According to Liljegren (2016: 90), Palula has a complex system of demonstratives with a 3-way contrast for proximate, distal, and objects that are out of sight (see example 62). Even though there were no examples for a visibility contrast in the field data, we can conclude that Palula has a distance- and visibility contrast for demonstratives.
Palula demonstratives, according to Liljegren (2016: 130).

<table>
<thead>
<tr>
<th></th>
<th>NOM, M, SG</th>
<th>non-NOM/PL/F</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROX</td>
<td>anu, enu</td>
<td>ani, eni</td>
</tr>
<tr>
<td>DIST</td>
<td>eɽo</td>
<td>eɽe</td>
</tr>
<tr>
<td>REM</td>
<td>eɽo</td>
<td>eɽe</td>
</tr>
</tbody>
</table>

For Kundal Shahi, the data was again rather difficult to interpret, as many sentences are missing. However, [aj] was used when referring to proximate objects, and [a:] for the objects that were a bit further away. These results are confirmed by Baart and Rehman (2005: 14). According to them, Kundal Shahi has a 3-way contrast in proximate, distal (within sight) and remote (invisible).

5.2.5 Summary

A variety of phonological, morphological, syntactic and lexical features were analyzed for the 9 Shina languages Brokskat, Kalkoti, Kohistani Shina, Kundal Shahi, Gilgiti Shina, Gurezi Shina, Palula, and Ushojo. A summary of the results is shown in Table 14.
Table 14: Summary of the feature analysis.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Brokskat</th>
<th>Kalkoti</th>
<th>Kohistani Shina</th>
<th>Kundal Shahi</th>
<th>Gilgiti Shina</th>
<th>Gurezi Shina</th>
<th>Palula</th>
<th>Sawi</th>
<th>Ushojo</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex syllable structure</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Mod. compl. syllable structure</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Voiceless aspirated consonants</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Voiced aspirated consonants</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Retroflex plosives</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Voiceless retroflex fricatives</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Voiced retroflex fricatives</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td><strong>Morph.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammatical gender</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Sex-based gender</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Ergative alignment</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Tripartite alignment</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td><strong>Synt.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOV word order</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Postpositions</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Lexicon</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numeral base 10</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Numeral base 20</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Numeral composition x+10</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Numeral composition 20+x</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>?</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Numeral composition x+20</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>?</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Lexicalization of kinship line contrast</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Lexicalization of direction contrast in kinship terms</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Lexicalization of age contrast in kinship terms</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>2-way distance contrast for demonstratives</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>3-way (or more) contrast for demonstratives</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>
6 Discussion

In 6.1.1, the results of the cognate analysis are discussed in relation to hypotheses about the relatedness of the languages from previous studies. In 6.1.2, general patterns in the structural analysis, as well as the results for particular features, are discussed. Areal patterns and other possible clusters are also reviewed. Finally, in 6.2, problems and limitations of the method are reviewed.

6.1 Discussion of Results

6.1.1 Cognate Analysis

The results of the cognate analysis were visualized with the software SplitsTree (Huson and Bryant, 2006), as shown in Figure 5. The corresponding distance matrix is already given in section 5.1 (Figure 4). The data suggests that all 9 languages are rather similar (average distance = 0.351, stdev = 0.033). However, what should be considered when interpreting the results is that they might have been influenced by misunderstandings of the questionnaire (even though the informants were instructed thoroughly before filling in the questionnaires), especially as there was only one informant per language. These issues are discussed in 6.2 again.

Figure 5: Visualization of the results from the cognate analysis (Created with SplitsTree 4; Huson and Bryant, 2006).

The SplitsTree visualization (Figure 5) shows that Gilgit Shina and Gurezi Shina are the languages that are the most similar to each other. They grouped together with Kohistani Shina and Brokskat in a cluster of the somewhat central Shina languages. The fact that Brokskat also belongs to this cluster confirms the hypothesis of Strand (2001: 253), that Brokskat is a descendant from Gilgit Shina.
The western Shina languages Sawi, Palula, and Kalkoti did also cluster together, however, they were not as similar as Gilgiti Shina and Gurezi Shina were to each other. This also confirms that these languages must have separated from the rest of the Shina languages earlier, as Liljegren (2009: 57-58) found as well.

Interestingly, Ushojo and Kohistani Shina are very far away from each other - even though Decker (1992: 67-68,70) proposed that they are closely related. Considering that the cognate analysis in this study was based on basic vocabulary, it is likely that the languages are not as closely related as previously assumed - however, in order to confirm this suggestion, a more thorough analysis would be needed, which would exceed the scope of this study.

Furthermore, Kundal Shahi stands out completely - which confirms its classification as an outlier. However, the results cannot tell us more about its origin either.

6.1.2 Feature Analysis

General Patterns

The feature analysis showed that the Shina languages do have very similar phonological, morphological, syntactic and lexical structures. Most features that appear typical for the Shina languages are also typical for the HKIA languages in general, such as SOV word order, the usage of postpositions, and sex-based grammatical gender (Liljegren, 2017: 131,139-140). Only few languages differed from the rest of the sample in some cases, and certain features showed more variation than others. The language that stood out the most is Kundal Shahi, which is interesting as it also is the language that is the most endangered and likely to be replaced by languages with a higher status in the region where it is spoken, namely Hindko (Baart and Rehman, 2005: 3).

Features with less variation

The syntactic and morphological features showed the least variation, and were consistent for most of the Shina languages. All 9 languages have SOV word order, use postpositions (even though case marking seems to be preferred in some cases), and have sex-based grammatical gender. Most Shina languages, except for Gurezi Shina and Sawi, also showed ergative-absolutive case alignment in noun phrases. Gurezi Shina and Sawi additionally mark the difference between S and O - meaning that they have a tripartite system. In the case of Sawi, that could be an areal influence from the neighbouring Gawarbati language, which also has tripartite case alignment (Liljegren, 2017: 133). In a broader perspective, there could also be an areal influence from languages with overt object marking, which are more common in the western part of the Hindu Kush region (Liljegren, 2014: 168).

Features with more variation

The phonological and lexical features showed more variation, which was not unexpected, as phonology in general tends to be influenced a lot by surrounding languages, especially in loanwords (Matras, 2007: 36-40). The variation in the lexical patterns can also be due to problems in the method, which will be further elaborated in 6.2.

The analysis of syllable structures showed a lot of variation, as the languages have both moderately complex and complex syllable structures. Figure 6 illustrates the geographical distribution of the different structures. At first, there are no clear recognizable patterns, and therefore it is likely to assume that contact with other non-Shina languages has induced the variation. However, Liljegren (2017: 124) found that IA languages in the center of the Hindu Kush region tend to have more restricted consonant clusters, whereas languages in the peripheral areas tend to allow more kinds of consonant clusters. These results are partially true for the Shina languages as well - Kohistani Shina, Gilgiti Shina and Gurezi Shina, which are all spoken rather central, have moderately complex syllable structures, whereas e.g. Palula, Kundal Shahi and Brokskat have complex syllable structures. Kalkoti and Sawi, even though classified as moderately complex, allow more elaborated consonant clusters than e.g. Gilgiti Shina,
Gurezi Shina or Kohistani Shina. An interesting remark here is also that e.g. the consonant clusters in Brokskat mainly appear in borrowings from Tibetan languages, such as [ʃr̥t͡ʃaːmæ] ‘tree’ in Brokskat and [ɫʃaŋma] ‘tree’ in Ladakhi (lbj) whereas e.g. Palula has retained some OIA clusters, as in [aːʂʈ] ‘eight’, from OIA *aṣṭā (Turner, 1966: 41). However, in order to explore this topic further, a more precise historical analysis would be needed.

![Figure 6: Syllable structures of the Shina languages (Map data ©2019 Google, created: 24/4/2019)](image)

Typically, IA languages have a 4-way contrast between voiced, voiced-aspirated, voiceless, and voiceless-aspirated (Masica, 1991: 101), whereas HKIA languages usually only have a 3-way contrast between voiced, voiceless and voiceless-aspirated consonants (Liljegren, 2017: 120).

The analysis showed that this is also the case for most Shina languages, even though the classification was not always clear-cut. All Shina languages showed some traces of an underlying contrast for voiceless-aspirated and voiced-aspirated consonants, but in most cases these contrasts have been replaced by e.g. tone, clusters with /h/, or another consonant (e.g. [pʰ] → [f]) in Sawi. It was also possible to see that the languages seemed to be in different stages of this process: Sawi has for example already lost all kinds of aspiration contrast (even though Buddruss (1967) was still able to notice it in his study, 50 years ago). Palula, on the other hand, still has contrasts for both voiceless and voiced aspirated consonants, but even here it was possible to notice a tendency to replace the aspiration of voiced consonants with consonant + /h/ clusters. The rest of the languages in the sample seemed to be somewhere in between those two extremes, the aspiration contrasts for voiceless consonants where still somewhat clear, whereas the contrasts for the voiced consonants have mostly been replaced already. A look at the word for ‘daughter’ in some Shina languages illustrates the process very well: in Sawi, [diː] does not show any indication of tone at all, but in Gilgiti Shina and Kohistani Shina [diː], there seems to be a rising tone, and in Palula [dʰiː], both aspiration and the same kind of tone pattern occur. However, this analysis is by far not complete, and should be explored further in future studies.

All Shina languages have retroflex plosives, but regarding the retroflex fricatives there is more variation: even though most of the languages do have voiced- and voiceless retroflex fricatives, the western Shina languages Palula, Sawi and Kalkoti do only have the voiceless retroflex fricative [ʂ] (the phonemic status of [ʐ] is unclear, especially its distinctness from the retroflex affricate [ɖʐ]), and Kundal Shahi does not have any retroflex fricatives at all. As illustrated in Figure 7, there seems to be a clear areal pattern, only Kundal Shahi stands out, which might again be explained by influence of surrounding languages, as e.g. Hindko does not have retroflex fricatives either (Mullick, 2019: 153), Liljegren (2017: 117-118) found that retroflex fricatives do not seem to occur in the south-east of the Hindu Kush region.

---

which also confirms that an areal influence is the reason for the absence of retroflex fricatives in Kundal Shahi. Also, future studies should look more precisely at the allophonic variation of retroflex fricatives and retroflex affricates, as this has not been analyzed with great detail in the current study.

Figure 7: Presence of retroflex fricatives in the Shina languages (Map data ©2019 Google, created: 24/4/2019)

In the lexical structure, there was more variation than in any of the other categories. However, conclusions, especially regarding the structure of kinship terms, should be discussed carefully, as they are very sensitive to errors when only relying on the answers of one informant. This matter is discussed further in 6.2.

Except for Kundal Shahi, all Shina languages have the numeral base 20. Kundal Shahi has the numeral base 10. This can again be explained by areal influence: a look at the data sets for Kashimir and Hindko, which are spoken in the same area as Kundal Shahi, did show that they also seem to have the numeral base 10. The analysis of numeral composition did show more variation, but as it is illustrated in Figure 8, it was difficult to find any areal patterns, especially without considering neighbouring languages.

Figure 8: Numeral composition of the Shina languages (Map data ©2019 Google, created: 24/4/2019)
Regarding the kinship terms, Brokskat and Gilgiti Shina were the only ones that lexicalize kinship line, and an age contrast has only been found in Palula and Gilgiti Shina. The results for the analysis of direction of kinship were especially diverse. Again, it was difficult to find any areal patterns, or patterns that can be explained with genetic relatedness. Problems with the method for the analysis of kinship terms are discussed in 6.2.

The demonstratives in Shina showed patterns that were typical for HKIA languages: they typically have multi-degree deictic systems, usually a two-way or three-way contrast for demonstratives (Liljegren, 2013: 145). As for most of the lexical features, it is difficult to explain why they do have these patterns, and a more thorough analysis would be needed.

**SplitsTree Analysis**

The results of the feature analysis were also visualized with the software *SplitsTree* (Huson and Bryant, 2006). Figure 9 shows the *Neighbour Net*, and Table 15 shows the corresponding distance matrix. However, as only 14 features were analyzed (resulting in 23 binary feature-categories for the SplitsTree analysis), and none of the surrounding, non-Shina languages were considered, the results should be interpreted carefully. These problems are discussed further in section 6.2.

![Figure 9: SplitsTree visualization of the results of the feature analysis](Created with SplitsTree 4; Huson and Bryant, 2006)

<table>
<thead>
<tr>
<th></th>
<th>bkk</th>
<th>xka</th>
<th>plk</th>
<th>shd</th>
<th>phl</th>
<th>sdg</th>
<th>scl (p)</th>
<th>scl (i)</th>
<th>ush</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>bkk</strong></td>
<td>0</td>
<td>0.391</td>
<td>0.130</td>
<td>0.286</td>
<td>0.217</td>
<td>0.304</td>
<td>0.217</td>
<td>0.478</td>
<td>0.261</td>
</tr>
<tr>
<td><strong>xka</strong></td>
<td>0.391</td>
<td>0</td>
<td>0.261</td>
<td>0.333</td>
<td>0.261</td>
<td>0.261</td>
<td>0.348</td>
<td>0.174</td>
<td>0.130</td>
</tr>
<tr>
<td><strong>plk</strong></td>
<td>0.130</td>
<td>0.261</td>
<td>0</td>
<td>0.333</td>
<td>0.173</td>
<td>0.261</td>
<td>0.348</td>
<td>0.348</td>
<td>0.304</td>
</tr>
<tr>
<td><strong>shd</strong></td>
<td>0.286</td>
<td>0.333</td>
<td>0.333</td>
<td>0</td>
<td>0.524</td>
<td>0.524</td>
<td>0.238</td>
<td>0.524</td>
<td>0.286</td>
</tr>
<tr>
<td><strong>scl (p)</strong></td>
<td>0.217</td>
<td>0.261</td>
<td>0.174</td>
<td>0.524</td>
<td>0</td>
<td>0.174</td>
<td>0.348</td>
<td>0.348</td>
<td>0.304</td>
</tr>
<tr>
<td><strong>scl (i)</strong></td>
<td>0.304</td>
<td>0.261</td>
<td>0.174</td>
<td>0.524</td>
<td>0.174</td>
<td>0</td>
<td>0.435</td>
<td>0.174</td>
<td>0.304</td>
</tr>
<tr>
<td><strong>phl</strong></td>
<td>0.217</td>
<td>0.348</td>
<td>0.261</td>
<td>0.238</td>
<td>0.348</td>
<td>0.435</td>
<td>0</td>
<td>0.522</td>
<td>0.304</td>
</tr>
<tr>
<td><strong>sdg</strong></td>
<td>0.478</td>
<td>0.173</td>
<td>0.348</td>
<td>0.524</td>
<td>0.348</td>
<td>0.174</td>
<td>0.522</td>
<td>0</td>
<td>0.304</td>
</tr>
<tr>
<td><strong>ush</strong></td>
<td>0.261</td>
<td>0.130</td>
<td>0.304</td>
<td>0.286</td>
<td>0.304</td>
<td>0.304</td>
<td>0.304</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Kundal Shahi and Palula are the languages that stand out the most, which might be explained by the fact that they both are spoken in rather isolated areas. Also, Palula has a rather high status in the villages where it is spoken (Liljegren, 2016: 23-25), making it less likely that the language has adopted patterns from other languages in the area, and rather has retained older structures, as e.g. in the case of the syllable structures and aspirated consonants. Kundal Shahi, on the other hand, is an endangered language, that is about to be replaced by other languages in the area (Baart and Rehman, 2005:3), making it likely that the language has already adopted many patterns from e.g. Hindko, another Indo-Aryan, non-Shina language.

The other two western Shina languages Sawi and Kalkoti were rather similar, and even Brokskat, Kohistani Shina, Gilgiti Shina, and Gurezi Shina, did form what could be an eastern cluster. However, Kohistani Shina and Ushojo did not cluster together, which was unexpected, considering that they are spoken close to each other, and have been hypothesized to be closely related to each other (Decker, 1992: 67,68-70). A more thorough analysis would be needed to confirm the validity of these findings.

6.2 Discussion of Method

Unfortunately, it was not possible to rely entirely on the field data for the analysis - instead it was necessary to use other data sources as well, which made the results less comparable than it was intended. Furthermore, only one informant per language was consulted, which also makes the data very prone to mistakes. This was especially noticeable in the kinship terms data, e.g. during the analysis of lexicalization of age contrast: Even though the informants were instructed carefully prior to filling in the questionnaires, it is possible that some of them simply gave a literal translation of ‘older brother’, ‘younger brother’ etc., instead of writing the actual lexicalized term they would usually use. Similar problems can have occurred in other questionnaires as well, e.g. the demonstrative questionnaire, where more examples from different speakers would have been needed as well. However, as the focus of the questionnaires was to compare a large sample of languages, these error sources had to be tolerated, but they should still be taken into account when interpreting the results.

Furthermore, the classification into binary values for the SplitsTree analysis was not always simple. Especially for the analysis of the aspiration contrasts, the values were not always yes or no - instead it seemed to be more of a continuous scale ranging from "no aspiration contrast at all" to "strong aspiration contrast".

As indicated in previous sections, by only looking at the Shina languages themselves, without considering any of the neighbouring languages, it was difficult to say much about possible areal patterns. Moreover, even though an attempt was made to choose features that were known to show variation in the HK region, they were still restricted by the availability of data. Analyzing a larger variety of features would have been helpful, but would unfortunately have exceeded the scope of this study.
7 Conclusions

The aim of this study was to create a micro-typological profile over 9 languages, that belong to the Shina branch of the HKIA languages. A cognate analysis was conducted, and a selection of phonological, morphological, syntactic, and lexical features was analyzed. The results were visualized with the software SplitsTree. In this sections the conclusions of the discussion, as well as the answers to the research questions are summarized. In addition, possibilities of future studies about the topic are discussed.

7.1 Answers to Research Questions

How do the Shina languages differ from each other lexically? Which languages stand out in some way?

The cognate analysis showed that all Shina languages are rather similar to each other, regarding their basic vocabulary. Kundal Shahi stood out the most, and even Ushojo did not cluster together with any other language. Sawi, Palula, and Kalkoti, however, clustered together, as well as Brokskat, Kohistani Shina, Gilgiti Shina, and Gurezi Shina.

Do the results from the cognate analysis confirm earlier hypotheses about the relatedness of the Shina languages among each other?

Yes, partially. The western Shina languages clearly form a cluster, confirming that they must have split from the "main" Shina languages a long time ago, as Liljegren (2009: 57-58) hypothesized in a previous study. Gilgiti Shina, Gurezi Shina, Brokskat and Kohistani Shina also form a cluster on their own, confirming that they belong to a more central cluster. An interesting remark here is also that Brokskat is included in this cluster as well, confirming that it also is a descendant from Gilgiti Shina, as hypothesized by Strand (2001: 253). Kohistani Shina and Ushojo, however, do not cluster together as it was expected, considering that Decker (1992: 67-68,70) proposed, on a basis of their common vocabulary, as well as the informants’ self-reported history, that they are closely related. Nonetheless, as this analysis is only preliminary, the question should be explored further before making any statements about it. Kundal Shahi stands out completely, confirming that it is an outlier, but also still making it difficult to classify and explain its origin.

Which features are shared by all Shina languages?

All Shina languages have moderately complex to complex syllable structures, SOV word order, postpositions, sex-based grammatical gender, and some form of distance contrast for demonstratives, which all are common for HKIA languages in general (Liljegren, 2017: 131,139-140,145). The data also suggests that ergative or tripartite case alignment in noun phrases is common for the Shina languages, as well as vigesimal numeral systems. Furthermore, the occurrence of retroflex plosives and fricatives, as well as an aspiration contrast for voiceless consonants, but not for voiced, was common.

Which features display more variation among the Shina languages?

Several features showed interesting patterns, and should be analyzed in future studies. The loss of aspiration for voiced (and in some cases even voiceless) consonants is such a feature - especially since there seems to be a connection with contrastive tonal patterns. These phonological changes, that the Shina languages, and probably many other HKIA languages as well, are undergoing, could be interesting to explore with more detail. Also, retroflex consonants, especially fricatives, should be analyzed in future studies as well. Furthermore, case alignment is something that could also be of great interest.
Most Shina languages showed ergative-absolutive alignment patterns in noun phrases, but also tripartite case alignment occurred. Here, the object marking could be of interest, as there seemed to be a tendency in e.g. Gurezi Shina to leave out the oblique case marker.

**How do the Shina languages differ from each other structurally? Which languages stand out in some way?**

In general, the languages were very similar to each other. The language that stood out the most was Kundal Shahi, which was the only language that has the numeral base 10, and also the only language that does not have any retroflex fricatives. Palula and Sawi also stood out phonologically, as Palula was the only language that still has preserved voiced and voiceless aspirates, and Sawi, on the other hand, has lost/ replaced those phonemes entirely.

**What can the results of the structural analysis tell us about any areal patterns?**

Several patterns that were found in the feature analysis can supposedly be explained by areal influence. The syllable structures, for instance, did show indications of areal influence, as well as the retroflex fricatives - although especially the latter should be explored more thoroughly before drawing further conclusions. The comparison of all languages in *SplitsTree* also showed indications of areal patterns, as Brokskat, Kohistani Shina, Gilgiti Shina, and Gurezi Shina formed an eastern cluster, and Sawi and Kalkoti, two western Shina languages, clustered together. Also, Kundal Shahi and Palula stood out as well - which in the case of Kundal Shahi can be explained by the fact that its speakers are shifting to speaking Hindko instead, and the language therefore might have adopted some patterns, and in the case of Palula, it can be explained by the language’s high status in the villages where it is spoken, making it likely that it has retained older features that other Shina languages already have lost.

**7.2 Future Studies**

In order to find out more about features which are unique for the Shina languages, whether there are any areal patterns, as well as how the languages are related to each other, a contrastive and historical analysis of the Shina languages, that also takes neighbouring languages into account, would be needed. This kind of study should also consider a larger variety of features.

As indicated earlier, features such as aspiration and retroflexion, as well as case alignment, seem to be of greater interest and should be taken up in future studies again. These analyses, however, would require data of several speakers, that mainly consists of free speech, and not only word lists and questionnaires.
References


A Figures and Tables

List of Tables

1. Other data sources that were used in the analysis ........................................... 11
2. The sample ........................................................................................................... 11
3. Elicitation sentences for analysis of demonstratives, based on Wilkins (1999) ... 19
4. Distance matrix of cognate analysis .................................................................. 21
5. Results of the syllable structure analysis ......................................................... 22
6. Results of the aspiration contrast analysis ...................................................... 25
7. Results of analysis of retroflexion ................................................................. 27
8. Results of grammatical gender analysis .......................................................... 29
9. Results of morphosyntactic alignment analysis .............................................. 31
10. Results of analysis of numeral bases ............................................................. 35
11. Results of analysis of numeral composition ................................................ 35
12. Results kinship terms analysis ....................................................................... 36
13. Results of demonstratives analysis .................................................................. 37
14. Summary of the feature analysis .................................................................... 41
15. Distance matrix of feature analysis ............................................................... 46

List of Figures

2. Classification of Shina languages according to Glottolog. (Hammarström et al., 2019) . 6
3. The Shina languages considered in this study. (Map data ©2019 Google, created: 8/4/2019) .... 12
4. Example of SplitsTree representation of results (Created with SplitsTree 4; Huson and Bryant, 2006) ................................................................. 19
5. Visualization of the results from the cognate analysis (Created with SplitsTree 4; Huson and Bryant, 2006) ................................................................. 42
9. SplitsTree visualization of the results of the feature analysis (Created with SplitsTree 4; Huson and Bryant, 2006) ................................................................. 46