A grammar sketch of Sauji
An Indo-Aryan language of Afghanistan

Nina Knobloch
A grammar sketch of Sauji

An Indo-Aryan language of Afghanistan

Sammanfattning

Denna studie presenterar ett urval av fonologiska och grammatiska drag i sauji, ett indoariskt språk som talas i en by i Kunarprovinsen i nordöstra Afghanistan. Sauji tillhör ett kluster av shinaspråk, som är en undergrupp av de hindukush-indoariska språken som talas i stora delar av nordligaste Pakistan, nord-östra Afghanistan och det omstridda Kashmirområdet. I likhet med många av språken i denna region är sauji knapphändigt beskrivet och därför är målet med den här studien att bidra med en grammatikskiss. Studien är baserat på data som har samlats in under fältarbete i regionen. Resultaten jämfördes med de närmast besläktade språken för att undersöka språket i en bredare kontext. Sauji är i stora drag väldigt likt palula, det närmast besläktade språket, men det har också visat sig att fonologin, lexikonet och även vissa grammatiska drag har påverkats mycket av gawarbati, ett annat indoariskt språk som talas i omgivningen.

Nyckelord

Sauji, shinaspråk, indoariska språk, Hindukush-indoariska, Afghanistan

Abstract

This study presents selected features in the phonology and grammar of Sauji, an Indo-Aryan language spoken in a village in the Kunar province in north-eastern Afghanistan. Sauji belongs to a cluster of (western) Shina languages - a subgroup of the Hindukush Indo-Aryan languages, which are spoken in large parts of northernmost Pakistan, north-eastern Afghanistan, and the disputed Kashmir region. As many languages in the Hindukush region, Sauji is largely underdescribed, hence the aim of this study was to provide a grammar sketch of the language, based on materials from field trips to the region. The results were compared to the closest related languages, to put the language into a broader context. Sauji is generally very similar to its closest linguistic relative, Palula, but also shows clear influence of Gawarbati, another Indo-Aryan language, on its phonology, lexicon, and some grammatical features.

Keywords

Sauji, Shina, Indo-Aryan languages, Hindukush Indo-Aryan, Afghanistan
## Contents

1. **Introduction** ......................................................... 2
   1.1 Aims and research questions .................................. 3
2. **Background** .......................................................... 4
   2.1 Language name and general setting ............................ 4
   2.2 Linguistic Setting .................................................. 5
      2.2.1 Genealogical and areal affiliation ......................... 5
      2.2.2 Language use and contact languages ....................... 6
   2.3 Previous research .................................................. 6
3. **Data and method** .................................................... 8
   3.1 Data ........................................................................ 8
   3.2 Analysis and procedure ........................................... 8
   3.3 Notes on transcription ............................................. 9
4. **Results** ................................................................. 10
   4.1 Typological overview .............................................. 10
   4.2 Phonology ............................................................. 12
      4.2.1 Consonant inventory ......................................... 12
      4.2.2 Vowel inventory ............................................... 14
      4.2.3 Phonotactics ..................................................... 15
   4.3 Nouns ..................................................................... 15
      4.3.1 Noun morphology ............................................... 15
      4.3.2 Gender .............................................................. 16
      4.3.3 Number ............................................................. 17
      4.3.4 Case ................................................................. 17
   4.4 Pronouns .................................................................. 19
      4.4.1 Personal pronouns .............................................. 20
      4.4.2 Demonstrative pronouns ..................................... 21
   4.5 Adjectives and quantifiers ......................................... 23
      4.5.1 Cardinal numbers .............................................. 23
   4.6 Postpositions .......................................................... 25
   4.7 Verbs ..................................................................... 27
      4.7.1 General properties ............................................. 27
      4.7.2 Verb morphology and conjugation classes ............... 27
      4.7.3 Tense, mood and aspect categories ....................... 29
   4.8 Grammatical relations .............................................. 31
      4.8.1 Verb agreement .................................................. 32
      4.8.2 Alignment of case marking of full noun phrases .......... 32
      4.8.3 Alignment of case marking of pronouns .................... 33
   4.9 Sentence modification ............................................... 34
      4.9.1 Interrogation ...................................................... 34
      4.9.2 Negation ........................................................... 35
5. **Discussion** ............................................................... 37
   5.1 Limitations and methodological issues ......................... 37
   5.2 Sauji in a broader typological context ......................... 37
      5.2.1 Areal and sub-areal influences ............................... 37
      5.2.2 Sauji compared to its closest linguistic relatives: Palula, Kalkoti and other Shina languages .......... 38
      5.2.3 Influences from next-door linguistic neighbours: Gawarbati and Pashto ....................... 39
   5.3 Changes since Buddruss’ (1967) field study ..................... 40
   5.4 Further research ...................................................... 40
Abbreviations

Glosses

1 First person
2 Second person
3 Third person
ABL Ablative case
ACC Accusative case
DAT Dative case
DET Determiner
ERG Ergative case
F Feminine grammatical gender
FUT Future tense
GEN Genitive case
IMP Imperative mood
IPFV Imperfective aspect
LOC Locative case
M Masculine grammatical gender
NEG Negation marker
NOM Nominative case
OBL Oblique case
PFV Perfective aspect
PL Plural
POSTP Postpositional case
PROX Proximate
PRS Present tense
PST Past tense
RED Reduplication
REM Remote
SG Singular
? Morpheme unknown/uncertain

Other abbreviations

A Subject in a transitive sentence
C Consonant
HKIA Hindukush Indo-Aryan
IA Indo-Aryan
P Object in a transitive sentence
S Subject in an intransitive sentence
V Vowel
1 Introduction

The aim of this thesis is to provide a description of selected topics in the phonology and grammar of Sauji, an Indo-Aryan language primarily spoken in Sau, a village located in the Hindu Kush mountains of northeastern Afghanistan. As many languages in the region, Sauji is generally very little described. Due to the war in Afghanistan many of its speakers were forced to relocate to refugee camps in Pakistan and it is uncertain how many of them have returned to the village, and to which extent they still use Sauji. The exact number of speakers is therefore difficult to determine.

The results presented in this study are based on first-hand data that was collected by Henrik Liljegren and/or Ajmal Nuristani during field trips to Jalalabad (February 2010), Kabul (April 2017), and Sau (July 2000) in Afghanistan, as well as two refugee camps in Pakistan (June 1999 - September 2000). It includes a number of word lists, free narratives, questionnaires, and text translations. A full index of the data is provided in Appendix B. The data was used to provide a description of a number of selected phonological and grammatical features of Sauji - partially to confirm the findings of Buddruss (1967) with more recent data, but also to fill in missing information that he has not been able to provide. The study is entirely based on field-data and therefore restricted by what the collected materials allow for. In the final parts of this thesis, Sauji is also compared to its closest relatives (Palula, Kalkoti and other Shina languages) and neighboring languages (e.g. Gawarbati and Pashto), in order to place the language in a broader areal-typological context, and attempt to explain the origin of certain features.

In order to give the reader more insight into the field data, a number of shorter, roughly glossed and translated, sample texts have been included in Appendix C.

Sauji has previously been described by Buddruss (1967) in a grammar sketch based on field data that he collected during a field trip in 1956. Furthermore, Liljegren (2009) provided a brief sketch of Sauji, mainly its phonology, in an article where he compared the language to its closest relatives Palula [phl] and Kalkoti [xka], traced their common proto-language, and linked Sauji historically to the southern Palula dialect. His article is partially based on Buddruss’ findings, as well as some a subset of the field data that was used in this study. However, besides these two publications, there is very little information about Sauji, which is why the aim of this thesis is to fill some of these gaps, as well as to clarify and confirm the previous findings.

The current study differs from Buddruss’ description in regards to data, method and theoretical foundation and does therefore provide a valuable addition to the body of knowledge about Sauji. First and foremost, the data used in this study is much more recent and comes from many different consultants. Buddruss (1967) was only able to work with one consultant, so his grammar sketch essentially only describes an ideolec. However, he had the advantage of being able to work directly with the consultant, whereas the current analysis is based on recordings and transcriptions done prior to the analysis done by the author. There is also a major difference in the theoretical approaches to the analysis of the different features - Buddruss’ description still has a much more European-centered perspective, which is for example noticeable in his description of Sauji noun morphology and TMA-categories. The current study is meant to provide a more modern, typologically informed perspective on Sauji. This also includes a comparison with closely related and geographically close languages, which is something that Buddruss has not been able to do in the same way, as many of these languages have only been documented and described in recent years. In addition, the broader data and different theoretical perspective have led to different analyses of certain features.
1.1 Aims and research questions

The work for this thesis has been guided by the following research questions:

1. What are the major phonological and grammatical features of Sauji? Focusing on:
   - Phonology: Phoneme inventory and syllable structure
   - Nouns: Noun morphology, gender, number and case
   - Pronouns: Personal and demonstrative pronouns
   - Adjectives and cardinal numbers
   - Postpositions
   - Verbs: Form and function of different TMA-categories
   - Grammatical relations: Verb agreement, alignment of case marking of full nouns phrases and pronouns
   - Sentence modification: Interrogation and negation

2. Which similarities and differences does Sauji show compared to...
   a) ... closely related languages, such as Palula and other Shina languages?
   b) ... its closest geographical neighbors, such as Gawarbati or Pashto?

3. Are there any recognizable changes that have happened in Sauji since Buddruss’ (1967) description?

The choice of features for question 1 was based on (1) the availability of sufficient data, (2) what had been described in previous work on Sauji and required confirmation or more attention, and (3) what had not been analyzed at all previously.
2 Background

This chapter provides an overview of the information that is already known about Sauji, starting with a discussion about the language name and the general setting, such as where and by whom the language is spoken, and how many speakers it still has in Section 2.1. In Section 2.2, the linguistic setting is outlined, which includes a discussion of the genealogical and areal affiliation of the language, as well as information about the language use and contact languages. In Section 2.3 a short overview of the few existing descriptions of the language is given.

Figure 1: Location of the Sau village in eastern Afghanistan, where Sauji is spoken. Created with ArcGIS online. Link to the map: https://arcg.is/1uP1Db.

2.1 Language name and general setting

Sauji (ISO 639-3: sdg, Glottocode: savi1242) is primarily spoken in Sau, a village in the Naray district of the Kunar province in eastern Afghanistan, close to the Kunar river and the border to Pakistan (see Figure 1). In previous publications the language has often been called Sawi or Savi, but as Decker (1992: 77-78) reported, its speakers refer to it as Sauji, which is why this name will be used throughout this thesis. Information about the extent to which the language is (and has been) spoken in the village is somewhat conflicting. In his work from 1941, Georg Morgenstierne estimated about 100 Sauji-speaking households (Morgenstierne, 1941: 7), but Kendall Decker’s consultants have reported 8000 - 12000 people living in the village before the war in Afghanistan (Decker, 1992: 78). Because of the war, many Sauji speakers had to move to refugee camps in Chitral and Dir in Pakistan (Decker, 1992: 78). In the Ethnologue Database, an estimate of 3000 speakers in 1983 is given (Simons and Fennig, 2020). One of the consultants for this study reported in 2010 that Sauji is still spoken by all inhabitants of the village, and there is even a small number of Sauji speaking households in Jalalabad and Kabul (25_S_100212).
Unfortunately, we do not know anything about the status of former refugees, how many there are and how many of them still use Sauji.

2.2 Linguistic Setting

2.2.1 Genealogical and areal affiliation

Sauji belongs to a subgroup of the Indo-Aryan languages that are mainly spoken in northern Pakistan, north-eastern Afghanistan, and Kashmir - the Hindu Kush Indo-Aryan languages (HKIA). For a long time, these languages have been referred to as Dardic languages, but this term has been recognized as somewhat controversial. Morgenstierne (1961: 139) pointed out that the languages in this region have simply retained some features and ‘missed’ some of the newer innovations in other Indo-Aryan languages, due to their isolation in the mountains - which he does not see as enough evidence to consider them as their own genealogical subgroup (see also Strand, 2001: 251, footnote 441). According to Bashir (2003: 822), it is not possible to explain the similarities within this group of languages exclusively with the Stammbaum model - geographic components play an important role as well. Strand (2001: 251) also rejects the usage of the term. Mascia (1991: 460) proposes that instead of focusing too much on a historical classification of these New Indo-Aryan (NIA) languages, it would be of greater interest to recognize a number of overlapping “genetic zones”. Following this view, Liljegren (2014: 135) has proposed the term “Hindu Kush Indo-Aryan” languages instead, which does not imply any particular genealogical classification, but still recognizes shared historical developments.

Figure 2 illustrates the classification of Sauji, considering both areal and generical clusters. The HKIA languages can be separated into 6 subgroups: Pashai, Kunar, Chitral, Kohistan, Shina, and Kashmiri (Bashir, 2003: 824-825). Sauji belongs to the Shina languages, more specifically to the cluster of western Shina languages, which also includes Palula [phl] and Kalkoti [xka] (Liljegren, 2009). As for most languages in the Hindu Kush region, the internal classification of the Shina languages is not entirely clear. In the Glottolog Database, which is based on a number of different language descriptions and classifications itself, the Shina languages (here called “Shinaic Languages”) are separated into (‘main’) Shina, Kohistanic Shina, and Western Shina, as well as the two outliers Kundal Shahi [shd] and Brokskat [bkk] (Hammarström et al., 2020). Radloff (1992: 103-113) recognizes four geographical clusters: a northern cluster (Gilgit valley), an eastern cluster (Astor, Baltistan, Dras), a western “Diamer cluster” (Chilas and other western
areas), and a “Kohistan cluster” (Jaltok, Palas, Kolai). Strand (2001: 253) argues for two main dialect centers - one in Chilas, and one in Gilgit - that all Shina languages originate from. During the long time of war and unrest in this particular region, many Shina speakers where forced to leave their original settlements, which is why the languages are scattered in the region today. Strand (2001: 253) also mentions a particular interest in displaced Shina-speaking communities in Swat, Dir Kohistan and Chitral, which is close to where Sauji is spoken today. Bashir (2003: 824-825) describes classification that is rather similar to Radloff’s - with a Kohistan cluster (overlapping with Radloff’s Diamer and Kohistan Cluster), an Astor cluster (equivalent to Radloff’s eastern cluster), and a Gilgit cluster (equivalent to Radloff’s northern cluster). She classifies Palula and Sauji as not belonging to any of these groups, and as Liljegren (2009, 2013) found later, the two languages form, together with Kalkoti, another Shina cluster.

According to Morgenstierne’s informants, the Palula speakers agreed that their ancestors have come from the Shina-speaking region of Chilas to Chitral 2-3 generations before the time of his expedition (Morgenstierne, 1941: 8). Buddruss (1967: 11) concluded from that, that the Sauji speakers most likely belonged to this group of people. Morgenstierne (1941: 9) recognized Sauji as simply a dialect of Palula, that has been influenced a lot by Gawarbati [gwt] due to the close contact of the two languages. Buddruss (1967: 11) reported that Sauji grammar and vocabulary are very similar to Palula, and that the two languages are mutually intelligible. However, when I played a recording of a Sauji text for a Palula speaker from Ashret, he mentioned that only some words seemed similar to Palula, especially the verbs, but he would not have understood the text completely without knowing beforehand what the story is about (Naseem Haider, personal communication). Decker (1992) found that the lexical similarity between Palula and Sauji is only 56-58%, which he assessed as “limited degree of comprehension”. Kalkoti, another close relative of Sauji, is a Shina language spoken in Kalkot, Dir Kohistan, Pakistan (Liljegren, 2009, 2013). It is lexically less similar to Sauji than Palula (Liljegren, 2009: 54), but also belongs to the cluster of western Shina languages.

2.2.2 Language use and contact languages

As one of the consultants for this study reported in 2010, Sauji has a strong status in the Sau village and is spoken by all inhabitants (25_S_100212). When outsiders come to the village, people will switch to a lingua franca, such as Pashto [pbi]. To my knowledge, there are no schools with teaching in Sauji. School is usually taught in Pashto, but Arabic, English, and Dari are subjects as well. An alphabet book, using a modified version of the Pashto script has been published recently (Muslim, 2017), but it is unclear to which extent it is used.

Despite the close relatedness of Palula and Sauji, there is not direct contact between the two communities anymore (Cacopardo and Cacopardo, 2001: 231-232). However, the people in Sau are, according to Liljegren (2009: 11), aware of the existence of the very similar “speech of Ashret”. Except for the language, the Sauji speakers are fully integrated into the surrounding Gawarbati speaking community (Cacopardo and Cacopardo, 2001: 231-232). Still, Gawarbati has influenced Sauji a significantly, especially the lexicon (Buddruss, 1967: 10-11).

2.3 Previous research

The only researcher to, prior to this study, collect data of Sauji in the Sau village itself was Wolfgang Lentz in 1935, during the German Hindukush-Expedition (Buddruss, 1967: 7). Unfortunately, he never published the materials, except for a short word list. Later on, the norwegian linguist Georg Morgenstierne gathered some information about Sauji as well. In his article about Palula, he mentions Sauji a few times, as one of his consultants was from Sau - however, unfortunately not L1 speaker of Sauji (Morgenstierne, 1941: 8-9).

Georg Buddruss then continued to work on the language in 1956 - although he never traveled to the Sau village, he was able to find a Sauji speaker in Jalalabad, Afghanistan (Buddruss, 1967: 7-11). Even though he was only able to work with the consultant for two days, he wrote a grammar sketch
of the language, that is of course not complete and only describes an ideolect, but still contains an
impressive amount of observations. Buddruss has described a variety of features in Sauji, including the
phoneme inventory, noun morphology, pronouns, postpositions, numerals < 20, verb morphology and
TMA-categories, subordination and word order. In addition, his description also includes a number of
transcribed and translated texts, and a word list.

Decker (1992) interviewed two men from Sau for a sociolinguistic survey of languages in Chitral in
1989. Henrik Liljegren, who has done a lot of work on Palula and other languages in the region, has
published some Sauji materials as well - e.g. an article about the relatedness of Sauji, Palula, and Kalkoti,
where an attempt to reconstruct aspects of their common source speech Proto-Dangari has been made,
and Sauji has been linked historically to the southern (Ashreti) Palula dialect (Liljegren, 2009).
3 Data and method

3.1 Data

The data for this study has been collected by Henrik Liljegren and/or Ajmal Nuristani during field trips to Jalalabad (February 2010), Kabul (April 2017), and Sau (July 2000) in Afghanistan, as well as two refugee camps in Pakistan (June 1999 - September 2000). During the most recent field trip in 2017, which was part of a multi-language data collection workshop held in Kabul, Najib-Ullah assisted with the recordings. The data is generally very diverse and consists mainly of recordings and transcriptions of word lists (mainly for phonological analysis), free narratives (including one with visual stimuli, the Pear Story), questionnaires, and text translations, as well as interview notes on the current language situation - in total from 14 different consultants.

In Appendix B, a full index of the different data sets is provided, which will also be used to refer to the source of examples and other information throughout this thesis. The index also provides more detailed information on the exact dates and locations of the data collection, as well as who has been involved in the different stages of data collection and processing. The recorded material has previously been annotated to a varying extent. Much of it has been roughly transcribed, and some of it translated, but a large part of it had to be processed specifically for this study. Many of the transcriptions, glosses and translations done by other people involved in the project were checked by the author prior to the analysis.

3.2 Analysis and procedure

The following features were investigated:

- Phonology: Phoneme inventory and syllable structure
- Nouns: Noun morphology, gender, number and case
- Pronouns: Personal and demonstrative pronouns
- Adjectives and cardinal numbers
- Postpositions
- Verbs: Form and function of different TMA-categories
- Grammatical relations: Verb agreement, alignment of case marking of full nouns phrases and pronouns
- Sentence modification: Interrogation and negation

The choice of features that are included in the grammar sketch was based on (1) the availability of sufficient data, (2) what had been described in previous work on Sauji and required confirmation/ more attention, and (3) what had not been analyzed at all previously. Many of the selected features have to some extent already been described by Buddruss (1967), such as the segmental phonology, noun morphology, personal and demonstrative pronouns, postpositions, or TMA-categories, but were assessed to benefit from further examination with the more recent data. Other features, such as grammatical relations, numerals >20 (and numeral composition), and sentence modification, were not included in Buddruss (1967) description and were therefore chosen for this grammar sketch. Overall, the goal was to include a broad variety of features to provide a representative overview of the phonology and grammar of Sauji that can be used for e.g. typological comparison. However, time constraints also had to be considered, which is why e.g. complex sentences were not included in this grammar sketch.
Theoretical literature about these features has been consulted prior to the analysis. These include e.g. many of the feature descriptions in the World Atlas of Language Structures (WALS), such as Consonant Inventories (Maddieson, 2013a), Vowel Quality Inventories (Maddieson, 2013c), Syllable Structure (Maddieson, 2013b), Numeral Bases (Comrie, 2013c), Alignment of Case Marking of Full Noun Phrases (Comrie, 2013a) - just to mention a few, as well as other commonly considered publications such as Dahl (1985) for the analysis of the tense and aspect forms.

A majority of the free narratives that were included in the data had already been roughly transcribed, translated and glossed prior to this study, but were checked and improved by the author. Questionnaires, such as the pronoun questionnaire (36_S_100213), sentence questionnaire (46_FS_100213), and valency questionnaire (51_FR_170405) had previously been transcribed, but needed to be checked and glossed by the author. Other data for e.g. the analysis of phonology did not have to be processed further by the author, as it was already well prepared. Parallel to the analysis, the transcribed texts and questionnaires were entered entered in FieldWorks Language Explorer (FLEx).

In the discussion part of the thesis (Chapter 5), Sauji is briefly compared to closely related languages (mainly Palula, but also other Shina languages), as well as its next-door linguistic neighbors (such as Gawarbati and Pashto), with the aim to place the language in a broader areal context, and to explain the origin of certain features. This comparison takes previous publications about these languages into account. The discussion is not aimed to be complete in any way, as this would have required a detailed historical analysis, which would have exceeded the scope of this thesis. Instead, it should be seen as a brief discussion of genealogical and areal influences on the language, which can give an indication of the origin of certain features of Sauji, considering that the language is in a rather special situation, being completely isolated from other Shina languages, but at the same time having very close contact with Gawarbati and other neighboring languages.

3.3 Notes on transcription

Except for the discussion of Sauji phonology, a simplified transcription is used in the examples throughout this thesis. This transcription system is commonly used in publications about Indo-Aryan languages and can be compared to a romanized orthographic representation of the language, considering the lack of an official writing system for Sauji. It is based on the transcriptions that Buddrus (1967) and Liljegren (2009) have used in their publications, with some smaller adaptations. Retroflex consonants are represented with a dot below the letter, aspirated consonants are followed by an h instead of a superscript h, and affricates are also represented differently. All remaining phonemes are represented as in IPA. An overview of the transcription used for each phoneme is provided in Section 4.2 after the phoneme inventory of Sauji has been presented.
4 Results

This chapter constitutes the main part of this thesis. Here, the results from the analysis of the different features in Sauji are presented, starting with a summary/typological overview in Section 4.1 and continuing with a more detailed account of each feature in the following sections.

4.1 Typological overview

**Phonology** Considering the criteria presented in the World Atlas of Language Structures, WALS, (Maddieson, 2013a), Sauji has a moderately large consonant inventory with at least 32 consonants (counting affricates as separate phonemes). Sauji has 8 vowel phonemes, of which only /a/ and /aː/ have a length contrast. In the other 7 vowel phonemes, length differences occur as well, but they are not contrastive and instead always combined with a quality difference. The syllable structure of Sauji would be classified as “moderately complex” in WALS (Maddieson, 2013b) - CV and CVC syllable are allowed, but syllables consisting of a single vowel can occur word-initial, and two-consonant clusters in the onset or the coda are restricted to certain combinations. The Sauji phoneme inventory and phonotactics are discussed in detail in Section 4.2.

**Nouns** Sauji nouns have inherent masculine or feminine gender, qualifying it as a sex-based grammatical gender, according to WALS (Corbett, 2013). Gender is usually assigned arbitrarily, except for a number of animate nouns where it is based on biological sex. For some (mostly, but not exclusively, animate) nouns, a feminine form can be derived from the masculine form. Grammatical gender is visible through e.g. verbs or adjectives that agree with the noun in gender and number. Nouns inflect for number and case. The case forms include the nominative, oblique, genitive, and an infrequently used ablative form. The nominative and oblique case are used for a number of different core functions, and in addition the oblique also fulfills other case functions and is for example used as a locative, and before postpositions. Regarding the case alignment of the core cases, Sauji follows an aspect-based split system. In constructions with verbs in the imperfective, A and S occur in the nominative, whereas in constructions with perfective verbs A is marked with the oblique, and S usually remains in the nominative. P can occur in either the nominative or the oblique, which most likely depends on definiteness, i.e. that definite objects are marked with the oblique, but indefinite objects remain in the nominative, which is a feature that has most likely developed under the influence of Gawarbati (Morgenstierne, 1950: 15). Another exception is that S can occasionally be marked with the oblique as well. This might however be lexically triggered, rather than syntactically, considering that is has been found for other languages in the HK region that there are certain intransitive verbs that require an ergative subject (Rehman, 2011). The morphological aspects of gender, number and case are discussed in Section 4.3 whereas case alignment if full noun phrases is discussed in Section 4.8.2.

**Pronouns** The first and second person personal pronouns in Sauji have a nominative, accusative and a syncretic ergative/genitive form. However, in the singular, the nominative and accusative are also expressed by the same form. The case alignment of these forms deviates slightly from full noun phrases, but is again based on an aspectual split. In the imperfective, A and S occur in the nominative, and P in the accusative, but in the perfective, there is a tripartite alignment instead, where A occurs in the ergative, S in the nominative, and P in the accusative. The third person personal pronouns conflate with the demonstrative pronouns, which means that they have an additional two-way distance contrast between proximate and remote. They have the same case forms as the other personal pronouns, however, the ergative and genitive forms are not syncretic. Regarding the alignment of case marking, the third person pronouns follow the same pattern as the other personal pronouns, i.e. a nominative-accusative

---

1 The subject in a transitive sentence
2 The subject in an intransitive sentence
3 The direct object in a transitive sentence
pattern in the imperfective, and a tripartite pattern in the perfective. Section 4.4 treats the personal and demonstrative pronouns in more detail, and in Section 4.8.3, alignment of case marking of pronouns is discussed further.

**Adjectives and quantifiers**  Adjectives and quantifiers have only been discussed shortly in this grammar sketch. Both attributive and predicative adjectives agree with the head noun in gender and number. Adjectives are usually placed before the nouns, however, predicative adjectives occur in the position of the object instead, i.e. after the head noun/subject and before the noun. Numerals in Sauji are based on a vigesimal system and are also placed before the noun they modify. Adjectives and quantifiers are discussed in Section 4.5.

**Postpositions**  Sauji has a number of postpositions that are used to express a variety of spatial and temporal relations. They always occur with the oblique form of the noun, or a separate postpositional form of the pronoun. Postpositions are discussed in Section 4.6.

**Verbs**  As for many other languages in the region, aspect is more central for verbs in Sauji than tense. Each verb has an imperfective and perfective stem, which additional tense, gender and number suffixes are added to, in order to form the different tense and aspect categories. The verb agrees with either the subject or the object in gender and number, again following an aspect-based split system. Verbs with an imperfective stem follow accusative agreement, whereas verbs with a perfective stem follow ergative agreement. Even though the morphological patterns of the different tense and aspect categories was possible to determine, their function could not be analyzed with great detail. However, based on Buddruss’ (1967) description, as well as a comparison with similar forms in Palula, it was possible to draw some tentative conclusions. It appears that Sauji has two basic and very frequently used tense/aspect categories, *Present Imperfective* and *Past Perfective*. The other tense and aspect forms (except for the future) are then based on the basic forms, by adding an additional tense suffix to the stem. These forms have preliminary been labelled as *Past Imperfective*, *Perfect*, and *Pluperfect*. The morphological composition and function of different verb forms in Sauji is discussed further in Section 4.7 and verb agreement is treated in Section 4.8.1.

**Sentence modification**  Sauji has a number of interrogative pronouns and adverbs, which are typically placed before the verb. Polar questions have a typical question intonation and sometimes start with a question marker, however, the word order remains the same as in declarative sentences. Negation is expressed with a negation particle which is placed right before the verb. Sentence modification is discussed in Section 4.9.

**Word order**  Even though word order is not discussed in great detail in this thesis, it might be relevant to mention a few patterns in this typological overview. The basic word order of Sauji is SOV, which is very typical considering the areal distribution in the world-wide sample presented in WALS (Dryer 2013), but also in the Hindu Kush region specifically (Liljegren 2017: 139). As previously mentioned, numerals and adjectives are placed before the noun they modify, but adpositions are placed after the noun.
4.2 Phonology

In this section, the consonant and vowel inventory of Sauji are presented, mainly based on previous descriptions by Buddrus (1967) and Liljegren (2009), but complemented with examples from the present data set.

4.2.1 Consonant inventory

The consonant inventory of Sauji has already been analyzed in previous publications (Buddrus 1967; Liljegren 2009), and will therefore not be discussed with great detail here. The inventory is shown in Table 1. The phonemes given in parenthesis did not occur in the current data, and even Buddrus (1967) and Liljegren (2009) have already expressed doubts about their status as phonemes (see also the short discussions about aspiration and loanwords in the following paragraphs). Examples of the occurrences of each phoneme in different positions within the word are given in Table 2.

Table 1: Sauji consonant phonemes. (based on analysis by Buddrus, 1967: 15 and Liljegren, 2009: 31)

<table>
<thead>
<tr>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental/Alveolar</th>
<th>Post-alveolar</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>p b t d ʈ ɖ k g q</td>
<td>tʰ (dʰ) (ʈʰ) kʰ (gʰ)</td>
<td>m</td>
<td>n</td>
<td>η</td>
<td>(ŋ)</td>
<td>(ŋ)</td>
<td>(ŋ)</td>
</tr>
<tr>
<td>Nasal</td>
<td>m n ɳ (ŋ)</td>
<td>r</td>
<td>t</td>
<td>x</td>
<td>y</td>
<td>h</td>
<td>(ŋ)</td>
<td>(ŋ)</td>
</tr>
<tr>
<td>Trill</td>
<td>(f) v s z</td>
<td>j</td>
<td>dʒ</td>
<td>(fʒ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>(f) v s z</td>
<td>j</td>
<td>dʒ</td>
<td>(fʒ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affricate</td>
<td>tʃ tʃʰ ʈʃ ʈʃʰ</td>
<td>dʒ</td>
<td>dʒ</td>
<td>(fʃ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral fricative</td>
<td>ɬ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral approximant</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Aspiration Both Buddrus (1967: 12-17) and Liljegren (2009: 30-32) discuss the phonemic relevance of aspiration. In the current data, there were generally very few aspirated consonants, which were exclusively voiceless consonants, and often only very weakly aspirated. In IA languages, an aspiration contrast for both voiceless and voiced consonants is common (Masica, 1991: 101), however, in HKIA languages this contrast is often only present for voiceless consonants (Liljegren, 2017: 119-120) - which is also the case for most Shina languages (Knobloch, 2019: 24-27). Buddrus (1967: 15) already noted that he was not entirely sure about the aspiration of the affricates (/tʃʰ/, /ʈʃʰ/) and voiced plosives (/bʰ/, /dʰ/, /gʰ/). In the more recent data, almost none of the examples that he gave for these phonemes were aspirated - e.g. bhyən 'sister' (Buddrus, 1967: 88) occurred as /beːɳ/ (50_FR_170404). (See also Liljegren (2009: 31) for more examples.) The only clear instance of an aspirated affricate was /tʃʰɑːlɛ/ 'goat'. For the voiceless aspirated plosives, however, the aspiration was still rather clear in most cases, as Buddrus (1967: 15) noted as well. As the examples in Table 2 show, the aspirated consonants only occur word-initial, except for two word-final cases.

Loanwords According to Buddrus (1967: 16), /x/, /ŋ/ and /f/ only occur in loanwords. /z/ and /ʃ/ occur quite frequently in the more recent data (see Table 2 for examples), but the other phonemes are less frequent or do not occur at all. Buddrus (1967: 116) only gives one example of /musəfeɾi/ ‘journey’, but in the current data this f is realized as pʰ instead. This relates to another observation of Buddrus (1967), that /pʰ/ was sometimes also realized as [φ] by his consultant. It might then be possible to conclude that words with an original /pʰ/ (loanwords or not) are reinterpreted as /f/ by some speakers - which also appears to be a development in Gawarbati, as Liljegren (2017: 120) noted.
Table 2: Sauji consonant phonemes with examples.

<table>
<thead>
<tr>
<th>Cons.</th>
<th>Initial</th>
<th>Medial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>paːɬu 'leaf'</td>
<td>paːpi ‘paternal aunt’</td>
<td>gap ‘word’</td>
</tr>
<tr>
<td>b</td>
<td>baːbu ‘father’</td>
<td>baːbu ‘father’</td>
<td>dįib ‘tongue’</td>
</tr>
<tr>
<td>t</td>
<td>tuː ‘you’</td>
<td>rate: ‘on’</td>
<td>rat ‘blood’</td>
</tr>
<tr>
<td>ṭ</td>
<td>tʰaːnu ‘do (simple present)’</td>
<td></td>
<td>haːt ‘hand’</td>
</tr>
<tr>
<td>d</td>
<td>duː ‘two’</td>
<td>daːnda ‘teeth’</td>
<td>daːnd ‘tooth’</td>
</tr>
<tr>
<td>ŋ</td>
<td>ŋaː ‘old’</td>
<td>baːta ‘stones’</td>
<td>baːt ‘stone’</td>
</tr>
<tr>
<td>k</td>
<td>kaːŋ ‘ear’</td>
<td>jakaːf ‘11’</td>
<td>jak ‘1’</td>
</tr>
<tr>
<td>ṭk²</td>
<td>ṭkʰ ‘eat’</td>
<td></td>
<td>naːkʰ ‘fingernail’</td>
</tr>
<tr>
<td>g</td>
<td>gaːnu ‘big’</td>
<td>hangaːru ‘fire’</td>
<td>jang ‘fight, conflict’</td>
</tr>
<tr>
<td>q</td>
<td>qaziː ‘judge’</td>
<td>kamaːqal ‘stupid’</td>
<td>qafuq ‘spoon’</td>
</tr>
<tr>
<td>ŋs²</td>
<td>ŋs ‘milk’</td>
<td>anfįː ‘eye’</td>
<td>infi ‘bear’</td>
</tr>
<tr>
<td>ŋf²</td>
<td>ŋf ‘breast’</td>
<td>kufﬁːro ‘dog’</td>
<td>tkuﬁ ‘scorpion’</td>
</tr>
<tr>
<td>ɖʒ²</td>
<td>ɖʒ ‘goat’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>q</td>
<td>qu ‘louse’</td>
<td>buːdʒiloː ‘hear (prs.msg)’</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>suː ‘sun’</td>
<td>asi ‘1PL.ERG’</td>
<td>naː ‘nose’</td>
</tr>
<tr>
<td>z</td>
<td>zindagi ‘life’</td>
<td>qazi: ‘judge’ (Buddruss 1967: 125)</td>
<td>me ‘table’</td>
</tr>
<tr>
<td>ŋs</td>
<td>ŋs ‘horn’</td>
<td>deːsɔ: ‘see (simple past)’</td>
<td>bariš ‘year’</td>
</tr>
<tr>
<td>s</td>
<td>ʃaŋko ‘wood’</td>
<td>uːfi ‘wind’</td>
<td>bi ‘twenty’</td>
</tr>
<tr>
<td>x</td>
<td>xalaka ‘people’</td>
<td>vaxti ‘time’</td>
<td>xwax ‘good’</td>
</tr>
<tr>
<td>y</td>
<td>y ‘ ‘</td>
<td>pāyas ‘clean’</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>hino ‘be (prs.msg)’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>rate: ‘on’</td>
<td>tari ‘star’</td>
<td>fʃor ‘four’</td>
</tr>
<tr>
<td>ŋ</td>
<td>ŋ ‘ ‘</td>
<td>bowŋi ‘paternal uncle’</td>
<td>doŋ ‘yesterday’</td>
</tr>
<tr>
<td>ɬ</td>
<td>ɬa: ‘three’</td>
<td>paːɬu ‘leaf’</td>
<td>gaɬ ‘skin’</td>
</tr>
<tr>
<td>ɬ</td>
<td>ɬa: ‘this’</td>
<td>alo ‘be (pst.msg)’</td>
<td>bal ‘good’</td>
</tr>
<tr>
<td>v</td>
<td>vaxti ‘time’</td>
<td>laːvu ‘child/small’</td>
<td>jandraːv ‘snake’</td>
</tr>
<tr>
<td>j</td>
<td>jinu: ‘liver’</td>
<td>dija ‘both’</td>
<td>jej ‘mother’</td>
</tr>
<tr>
<td>m</td>
<td>mɪʃa ‘husband’</td>
<td>kirmi ‘worm’</td>
<td>nɛm ‘name’</td>
</tr>
<tr>
<td>n</td>
<td>nas ‘nose’</td>
<td>jinu: ‘liver’</td>
<td>ban ‘close’</td>
</tr>
<tr>
<td>ŋ</td>
<td>ŋ ‘ ‘</td>
<td>anдо: ‘egg’</td>
<td>beŋ ‘sister’</td>
</tr>
</tbody>
</table>

As already indicated in Section 3.3, a different transcription than IPA will be used for the examples throughout this thesis. The transcription system is presented in Table 3.

### Table 3: Transcription of Sauji consonant phonemes in this thesis.

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labio-</th>
<th>Dental/</th>
<th>Alveolar</th>
<th>Post-</th>
<th>alveolar</th>
<th>Retro-</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plosive</strong></td>
<td>p</td>
<td>b</td>
<td>t</td>
<td>d</td>
<td>t̪</td>
<td>d̪</td>
<td>k</td>
<td>g</td>
<td>q</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nasal</strong></td>
<td>m</td>
<td>n</td>
<td>n̄</td>
<td>(n̄)</td>
<td>(n̄)</td>
<td>(n̄)</td>
<td>(n̄)</td>
<td>(n̄)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Trill</strong></td>
<td>(f̌)</td>
<td>r</td>
<td>ř</td>
<td></td>
<td>ř</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fricative</strong></td>
<td>(⁷)</td>
<td>(w̌)</td>
<td>š</td>
<td>(x̌)</td>
<td>š</td>
<td>x̌</td>
<td>y̌</td>
<td>ȟ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Affricate</strong></td>
<td>(tš)</td>
<td>(č)</td>
<td>ǰ</td>
<td>(cȟ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lateral fricative</strong></td>
<td>ľ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Approximant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lateral approximant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2.2 Vowel inventory

Figure 3 shows the Sauji vowel system as both Buddruss (1967: 12-13) and Liljegren (2009: 32) have analyzed it. Table 4 shows the phonemes with their respective phonetic realizations as provided by Liljegren (2009: 32), as well as their representation in Buddruss’ (1967) description. It also includes broadly transcribed examples from the current data set, as well as the simplified transcription that is used in the examples throughout this thesis.

Both Buddruss (1967) and Liljegren (2009) discussed the relevance of vowel length in Sauji. Buddruss (1967: 12) only analyzed /a/ as having a phonemically relevant length contrast, the other vowels that show some kind of length contrast do also have a difference in quality (such as /e/ and /ɛ/), which is why he does not interpret length as relevant here. Buddruss (1967: 13) also mentions that /o/ often alternates with /u/ (not only word final), and /ɔ/ sometimes alternates with /a/ - it is therefore not entirely clear whether they actually are separate phonemes, but the vowel system that Buddruss (1967: 12-13) proposed is based on the hypothesis that they are. Henceforth, I will follow his example in this thesis and treat them as separate phonemes, even though I also noticed some difficulties in distinguishing clearly between /a/ and /ɔ/, as well as /o/ and /u/. Especially /o/ and /u/ occur quite often in the same places, for example the (masculine singular) Past Perfective form ending on -aalu and -aaloo. Here, a clue to this variation might be stress: when /u/ occurs the stress of the word was on the previous syllable, whereas when /o/ occurs, the stress was on the syllable with this vowel instead.

### Table 4: Sauji vowel phoneme inventory with examples.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>[i]/[iː]/[a]</td>
<td>i</td>
<td>/iːnuː/ 'liver', /miːaː/ 'husband'</td>
<td>i</td>
</tr>
<tr>
<td>e</td>
<td>[e]/[eː]</td>
<td>e</td>
<td>/beː/ 'we', /hineː/ 'be (mpl)'</td>
<td>ee</td>
</tr>
<tr>
<td>ɛ</td>
<td>[ɛ]/[a]</td>
<td>ɛ</td>
<td>/deː/ 'ten', /hineː/ 'be (fpl)'</td>
<td>e</td>
</tr>
<tr>
<td>a</td>
<td>[a]/[aː]</td>
<td>a</td>
<td>/jak/ 'one', /maːʃin/ 'fish'</td>
<td>a</td>
</tr>
<tr>
<td>aː</td>
<td>[ɑ]/[aː]</td>
<td>ā</td>
<td>/tariː/ 'star', /kitaː/ 'book'</td>
<td>aa</td>
</tr>
<tr>
<td>ɔ</td>
<td>[ɔ]/[ɔː]/[o]/[oː]</td>
<td>ɔ</td>
<td>/soː/ 'six', /bowrɛː/ 'paternal uncle'</td>
<td>o</td>
</tr>
<tr>
<td>o</td>
<td>[o]/[oː]</td>
<td>o</td>
<td>/goːʃ/ 'house', /fʃɔːr/ 'four'</td>
<td>oo</td>
</tr>
<tr>
<td>u</td>
<td>[u]/[uː]</td>
<td>u</td>
<td>/surriː/ 'sun', /ʃuːrɔː/ 'breast'</td>
<td>u</td>
</tr>
</tbody>
</table>

Figure 3: Sauji vowel phoneme inventory. (based on analysis by Buddruss 1967: 12-13 and Liljegren 2009: 32)
4.2.3 Phonotactics

Sauji allows for open (CV) and closed syllables (CVC). Two-consonant clusters are allowed as well, however, in the onset only with /j/ or /ɾ/, and in the coda only with a nasal. Syllables consisting of a single vowel also occur word-initially. Examples for all the possible structures are provided in (1).

(1) a. CV - /haː.ɖu/ ‘bone’
   b. CVC - /daːn.du/ ‘tooth’
   c. CCV(C) - /braː/ ‘brother’, /graːm/ ‘village’
   d. CVCC - /pʰaːnt/ ‘path’
   e. V - /uː.ʃiː/ ‘wind’

Aspirated consonants mostly occur word-initial (with few exceptions). Furthermore, voiced plosives and affricates seldom occur word-final. However, this appears to vary from speaker to speaker: some consultants used them word finally (such as /haːɖ/ ‘bone’ or /daːnd/ ‘tooth’), whereas another one preferred to add a vowel in the end (such as /haː.ɖu/ ‘bone’ or /daːn.du/ ‘tooth’). Moreover, /h/ only occurred word-initially, and /ŋ/ and /ɾ/ only medial and final.

Stress and pitch accent have not been analyzed specifically in this thesis, but considering that they are phonologically relevant features in e.g. Palula (Liljegren, 2016: 72-75) and other Shina languages, they might also be of relevance for Sauji.

4.3 Nouns

In Sauji, nouns have inherent masculine or feminine gender. They function syntactically as the head of noun phrases. Within the noun phrase, they appear to always occur finally (as examples (2a-c show), except for relative clauses, where the head noun is placed before the relative clause, according to Buddrus (1967: 59) examples. Adjectives agree with the head noun in gender and number. In this section, the main focus lies on the morphological marking of number and case, and how grammatical gender is reflected in Sauji.

(2) a. leew-i pu-yee
   little-f.sg girl-obl
   ‘Little girl’
   b. čur yon-e
   four month-pl
   ‘Four months’
   c. le xalaka
   det.prox.pl.nom people
   ‘These people’

4.3.1 Noun morphology

The nominative singular forms in Sauji are usually not morphologically marked. However, plural and case are marked with a number of inflectional suffixes. Grammatical gender (masculine or feminine) is inherent and visible through adjective or verb agreement. Nonetheless, some feminine nouns can be derived from their masculine counterparts. For most nouns, the stem that inflectional suffixes are added to, corresponds to the root of the noun. However, for the feminine nouns that were derived from their masculine counterparts, the stem consists of the root and the derivational suffix.
Inflectional morphology and inflectional classes  Number and case in Sauji are expressed by means of inflectional morphology, i.e. a number of different suffixes that are added to the stem. In masculine forms that end in -u/-oo, the last vowel is always deleted before the plural or case suffix. However, in feminine forms (often the ones that are derived from another masculine form) that end in -i, the last vowel is removed before adding the plural suffix, but remains when adding a case suffix.

Based on the nominative singular forms and plural inflections, the nouns can be sorted into three inflectional classes. The first one consists of all nouns that end in a consonant in the nominative singular, as well as nouns that end in -i (which appear to always be feminine and monosyllabic), but where -i belongs to the root of the word. Here the whole nominative form constitutes the stem, and plural and case suffixes are added to it accordingly. Some examples are provided in 3. The second class consists of exclusively masculine nouns that end in -u/-oo, such as the words provided in example 4. Here, the last vowel is deleted before adding a case or plural suffix. In nouns that belong to the third group, i.e. feminine forms that end in -i, which are often (but not exclusively) derived from another masculine form, the last vowel is removed before adding the plural suffix, but remains when adding a case suffix. Two examples of these nouns are provided in 5. The particular plural and case inflections for each class are treated in sections 4.3.3 and 4.3.4.

(3) a. maanuṣ ‘man’  (4) a. kučuroo ‘dog’  (5) a. piši ‘cat’
b. si ‘bridge’  b. piloo ‘ant’  b. suri ‘sun’

Derivational morphology  Derivational morphology is less important for nouns in Sauji, however, there are some instances where feminine nouns are derived from their masculine counterparts. These examples are often based on biological sex (which is then reflected in the grammatical gender of the word) and include animate nouns such as words for animals, or kinship terms. For masculine nouns that end in -u/-oo, the last vowel is exchanged with -i, such as in kučuroo ‘male dog’ and kučuri ‘female dog’, or nawaasu ‘grandson’ and nawaasi ‘granddaughter’. When the masculine form ends on a consonant, the feminine form is derived by adding -i directly to this form, such as in kukur ‘male chicken’ and kukuri ‘female chicken’. In one example, a longer morpheme (-eṛi) was added: inč ‘male bear’ and inčeṛi ‘female bear’. Even though it has not been confirmed in the current data, it might also be possible to use this type of derivation as a diminutive form, which is for example the case in Palula (Liljegren, 2016: 99-100). Comparing for example the Sauji words anguri ‘finger’ and anguto ‘thumb’ (Buddruss, 1967: 80), there might be a similar pattern as in Palula.

4.3.2 Gender

As already indicated in the previous section, nouns in Sauji have inherent masculine or feminine gender, which is reflected in adjective- and verb agreement. This is illustrated in examples 6a and b. The agreement markers on adjectives and verbs are very consistently -i for the feminine forms, and -u/-o for the masculine.

(6) a. se gooshi subaanne-u hin-u
   det.rem.nom house beautiful-m.sg be.prs-m.sg
   ‘That house is beautiful.’ (56_FR_170403)
b. gooshi ee muškani ek nil-i drab hin-i [...] 
   house-obl in.front.of one green-f.sg grass be.prs-f.sg
   ‘In front of the house there is a green lawn [...]’ (14_HR_000519)

Except for the derived feminine forms (as explained in the paragraph about derivational morphology in Section 4.3.1), grammatical gender is not phonologically signalled on the noun itself. However, there are still some semantic and formal clues to gender assignment in Sauji. Animate nouns usually take the
grammatical gender based on their biological sex. Nouns that end in -u/-oo in the nominative singular, which are often animate, are always masculine. Nouns that end in -i (where -i is either a derivational morpheme or belongs to the root) are almost exclusively feminine, with few exceptions. However, nouns that end in a consonant (and are in most cases inanimate) can either be masculine or feminine and there are no formal or semantic clues to their grammatical gender.

4.3.3 Number

The plural inflections in the nominative are based on the inflectional classes that have been presented in Section 4.3.1. Sauji has three different plural suffixes: -a, -ee and -e. Nouns that belong to the first group (i.e. end in a consonant or non-suffixal -i) take the suffix -a, as illustrated in example 7.

(7) a. kitaab ‘book’ (m) - kitaaba ‘books’
   b. meez ‘table’ (m) - meezaa ‘tables’
   c. si ‘bridge’ (f) - sia ‘bridges’

The masculine nouns from the second group, i.e. those that end in -u/-oo in the nominative singular, take the plural suffix -ee, while deleting the final -u/-oo. Examples are presented in 8.

(8) a. kučuroo ‘male dog’ (m) - kučuree ‘dogs’
   b. piloo ‘ant’ (m) - pilee ‘ants’

Nouns from the third group, i.e. feminine nouns that end in -i in the nominative singular and are not monosyllabic, take the plural suffix -e instead. Again, the final vowel is deleted before adding the suffix. Examples are provided in 9.

(9) a. suri ‘sun’ (f) - sure ‘suns’
   b. anguri ‘finger’ (f) - angure ‘fingers’
   c. kukuri ‘hen’ (f) - kukure ‘hens’

The only exceptions to the presented patterns that occurred in the field data were bra ‘brother’, which is brawu in the plural, and pu ‘boy’, which is jitaka in the plural. According to Buddrus [1967:36], some feminine nouns ending in consonants take the plural suffix -e as well. However, I was not able to confirm this with the more recent data. Looking at some of the examples that he provided, and comparing them to the more recent data, the plural of ‘tongue’ jib is clearly jiba, and the plural of ‘sister’ been is clearly beena, which shows that they belong to the first group.

4.3.4 Case

Sauji has two core cases, the nominative and the oblique, as well as a genitive case (which often takes the same form as the oblique), and a very infrequently used ablative. Table 5 summarizes the different case suffixes in Sauji, which are usually added directly to the stem of the noun.

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>∅</td>
<td>-a/-e/-ee</td>
</tr>
<tr>
<td>Oblique</td>
<td>-(y)ee</td>
<td>-u (-o/-un)</td>
</tr>
<tr>
<td>Genitive</td>
<td>-(y)ee, -an</td>
<td>-u (-o/-un)</td>
</tr>
<tr>
<td>Ablative</td>
<td>-an</td>
<td>?</td>
</tr>
</tbody>
</table>

Table 5: Case suffixes in Sauji.
Core cases: nominative and oblique  The nominative singular is not marked morphologically, and the nominative plural suffixes have already been treated in the previous section (4.3.3). The oblique, is marked morphologically with the suffix -ee in the singular (-yee before vowels). The oblique plural is marked with the suffix -u. Buddruss (1967: 36) also mentions -o or -un, often nasalized. Example 10 shows the different nominative and oblique forms for the word maanus ‘man’ in Sauji.

(10) a. maanus  b. maanus-a  c. maanus-ee  d. maanus-u  
man(NOM.SG)  man-NOM.PL  man-OBL.SG  man-OBL.PL

There is also a small number of irregular and suppletive forms where the stems are changed as well, but they usually have the same case/number endings. Example 11 shows the nominative and oblique forms for the word pu ‘boy’. Another irregular form is teeri ‘woman’, which remains unchanged in the oblique singular.

(11) a. pu  b. jitak-a  c. pi-yee  d. jitak-u  
boy(NOM.SG)  boy-NOM.PL  boy-OBL.SG  boy-OBL.PL

The nominative is used for the subject of transitive sentences (A) in the imperfective, and for the subject of intransitive sentences (S), as examples 12 and 13 illustrate. It is also, in some contexts, used for the direct object in transitive sentences (P) - which is often also marked with the oblique. Examples of both uses are provided in 14. There is no entirely clear pattern in the field data, but Buddruss (1967: 33) mentions that this pattern of P occurring in either the nominative or the oblique (which he calls accusative) might have to do with definiteness - the oblique is used for definite forms, and the nominative for the indefinite. This matter is discussed further in Section 4.8.2.

(12) laaw-u  pu  busaag-u  (13) goos  dajaan-u
little-M.SG  boy(NOM)  hungry-M.SG  house(NOM)  burn.IPFV-M.SG
hin-u  be.PRS-M.SG

‘The little boy is hungry.’ (51_FR_170405)

(14) a. maanus-ee  inčeri  deeš-i
man-OBL  bear(F.NOM)  see.PFV-F.SG

‘The man saw the (female) bear.’ (51_FR_170405)

b. inčeri-yee  maanus-ee  biyaal-u
bear(F)-OBL  man-OBL  frighten.PFV-M.SG

‘The (female) bear frightened the man.’ (51_FR_170405)

Besides being used for the object of transitive sentences (P) (as illustrated in example 14b), the oblique is also used for the subject of transitive sentences (A) in the perfective - therefore functioning as ergative case as well. This is illustrated in example 15.

(15) pu-yee  pi-yee  piyaanil-u
girl-OBL  boy-OBL  know.PFV-M.SG

‘The girl knew the boy.’ (51_FR_170405)

4Here my analysis deviates slightly from the one done by Buddruss (1967: 31-35), which has mainly to do with the fact that he separated the different functions of what I summarized as "oblique". Summarizing them seemed the most fitting in this case, since all the suffixes for the accusative, ergative, postpositional functions (which will be explained in the following) etc. take the same form, and Buddruss (1967) did the same for the plural oblique forms.
Other cases and case functions  The oblique also fulfills a number of other case functions such as postpositional case, genitive, and locative, as illustrated in examples (16), (17), and (18) respectively. According to Buddruss (1967: 32), examples such as (16) would be classified as dative case instead, where the noun then takes the case suffix -(y)ee. However, based on the evidence in the current data, it appears to be possible to include these forms in the postpositional case with the suffix -(y)ee, which is then used together with postpositions such as (but not only) tee 'to, towards, at'. This matter is discussed further in Section 4.4 about postpositions.

Regarding the genitive case in the singular, besides the suffix -(y)ee (see example 17) that occurred in the current data, Buddruss (1967: 31) also found that -ã/-õ/-aan (which are similar to the ablative suffix) were used. However, he notes that -ee is mainly used for animate nouns, and -ã/-õ/-aan for inanimate, which would explain why the latter did not occur in the current data - there were only examples of genitive used with animate nouns.

(16)  pi-yee  poo-yee  tee  agil-u  
   boy-obl  girl-obl  to  look.pfv-m.sg
   ‘The boy looked at the girl.’  (51_FR_170405)

(17) a.  bowriyee  lerı  ‘paternal uncle’s wife’  (50_FR_170404)
   b.  saraaniyee  miš  ‘wife’s sisters husband’  (50_FR_170404)

(18)  se  dwi  goosh-ee  hin-i  
   3sg.rem.nom  other  room-obl  be.prs-f.sg
   ‘It [the cat] is in the other room.’  (56_S_100213)

Sauji also has an ablative case with the suffix -aan (or according to Buddruss (1967: 35) sometimes also a slightly nazalized [ã] or [õ]), which is used for motions away from something. An example is provided in (19).

(19)  ler-u  meez-aan  sum  paayas  til-u  
   woman-pl.obl  table-abl  soil  clean  do.pfv-m.sg
   ‘The women wiped dirt of the table.’  (51_FR_170405)

Buddruss (1967: 37) mentions that, at least in the plural, the oblique case is also used to express the instrumental, but unfortunately such examples were missing in the current data. In the singular, however, the consultants used postpositions to express instrumental meaning, as in example (20).

(20)  pi-yee  jandraaw-ee  ratee  šaank-ee  dee  sarsari  saat  dit-i  
   boy-obl  snake-obl  on  stick-obl  with  ?  ?  give.pfv-f.sg
   The boy hit the snake with a stick.  (51_FR_170405)

4.4 Pronouns

In this section, the personal and demonstrative pronouns in Sauji are presented and discussed. The third person personal pronouns are treated in Section 4.4.2 since they correspond to the demonstrative pronouns, and the first and second person personal pronouns are treated in Section 4.4.1.

5bowri ‘paternal uncle’ and saraani ‘wife’s sister’ in example (7) are lexicalized.
4.4.1 Personal pronouns

The first and second person personal pronouns in Sauji occur in the nominative, accusative and genitive/ergative case. An overview of the different forms is given in Table 6. In the singular, the nominative and accusative forms are the same, however, in the plural there is a difference. Furthermore, the second person singular and plural forms in the nominative are the same.

<table>
<thead>
<tr>
<th>Case</th>
<th>1SG</th>
<th>2SG</th>
<th>1PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>ma</td>
<td>tu</td>
<td>be</td>
<td>tu</td>
</tr>
<tr>
<td>Accusative</td>
<td>ma</td>
<td>tu</td>
<td>asoo</td>
<td>tusoo</td>
</tr>
<tr>
<td>Ergative/Genitive</td>
<td>mi</td>
<td>či</td>
<td>asi</td>
<td>tusi</td>
</tr>
<tr>
<td>Postpositional</td>
<td>ma</td>
<td>tu</td>
<td>ason</td>
<td>tusan</td>
</tr>
</tbody>
</table>

The nominative form is used for the subject of imperfective, transitive sentences (A), as well as for the subject of all intransitive sentences (S), as examples (21) and (22) show. The accusative is used for the object of all transitive sentences (P), as illustrated in example (23). In perfective sentences, the subject occurs in the ergative instead (see example (24)) - resulting in a tripartite alignment pattern in perfective constructions, and a nominative-accusative alignment in imperfective constructions. The alignment of case marking of pronouns is discussed further in Section 4.8.3. The ergative and genitive forms are the same, as examples (24) and (25) illustrate.

(21) tu lasee piyonaan-o
     2SG.NOM 3SG.PROX.ACC KNOW.PFV-M.SG
     ‘Do you (sg) know him?’ (36_S_100213)

(22) tu kanee beet-ee-n-ee
     2PL.NOM where sit.PFV-M.PL-PRS-M.PL
     ‘Where do you (pl) sit?’ (36_S_100213)

(23) door ta piyee ma dees-oo
     yesterday DET.REM.SG.NOM boy 1SG.ACC see.PFV-M.SG
     ‘Yesterday, that boy saw me.’ (36_S_100213)

(24) asi tasee moril-oo
     1PL.ERG 3SG.REM.ACC kill.PFV-M.SG
     ‘We killed him.’ (36_S_100213)

(25) či asi gooš dees-oo
     2SG.ERG 1PL.GEN house see.PFV-M.SG
     ‘Did you (sg) see our house?’ (36_S_100213)

The dative case forms that Buddruss (1967: 39) provided in his description, maṭee, tuṭee, asonṭee and tusoṇṭee, require some additional discussion. Buddruss (1967: 39-40) considers them dative forms, but also points out the similarities with the postposition tee ‘to’. In addition to the dative forms, he also
provides special postpositional case forms that are used in combination with the other postpositions: ma, tu, asan/asa and tusan/tusa (Buddruss, 1967: 39). What is noteworthy here is that he marks the last a in the plural forms as nasalized (which would look something like [asân], [asã], [tusân] and [tusã] in phonetic transcription). In the current data there were also forms such as asondiyo 'from us' - which includes the postposition diyo 'from'. This leads me to the conclusion that it would be better to analyze both case forms as a postpositional case, which is then used together with different postpositions, instead of having both a dative and a postpositional case. However, this certainly links to a larger discussion about dative case and postpositions, which will be continued in Section 4.6. Examples 26 and 27 show different analyses of the dative or postpositional constructions.

(26) ti la kitaab tufee dit-oo
3SG.REM.ERG DET.PROX.SG.OBL book 2SG.DAT give.PFV-M.SG
‘Did he give this book to you (sg)?’ (36_S_100213)

(27) ti la kitaab tu tfee dit-oo
3SG.REM.ERG DET.PROX.SG.OBL book 2SG.POSTP to give.PFV-M.SG
‘Did he give this book to you (sg)?’ (36_S_100213)

4.4.2 Demonstrative pronouns

As already mentioned in Section 4.4.1, the third person pronouns correspond to the demonstrative pronouns, which is why they are presented in this section instead. Unfortunately, the data did not contain enough examples to provide the full paradigm, which is why it had to be supplemented with some of Buddruss’ (1967) examples. An overview of the paradigm is presented in Table 7. The forms in parenthesis have not been confirmed in the field data, but have instead been taken from Buddruss (1967: 41-43). If two forms, one in parenthesis and one without are given, it means that Buddruss’ form (in parenthesis) deviates slightly from the one found in the field data. Otherwise, the analysis largely agrees with the one provided by Buddruss (1967). There is also another case form, the postpositional case, which will be discussed later in the text.

<table>
<thead>
<tr>
<th>Determiner</th>
<th>Pronoun</th>
<th>Determiner</th>
<th>Pronoun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>Plural</td>
<td>Singular</td>
<td>Plural</td>
</tr>
<tr>
<td>NOM</td>
<td>la</td>
<td>le</td>
<td>la</td>
</tr>
<tr>
<td>ACC</td>
<td>la</td>
<td>lena</td>
<td>lasee</td>
</tr>
<tr>
<td>ERG</td>
<td>li</td>
<td>lini (leni)</td>
<td>li</td>
</tr>
<tr>
<td>GEN</td>
<td>(lesi)</td>
<td>(leni)</td>
<td>(lesi)</td>
</tr>
</tbody>
</table>

In Sauji, the demonstrative pronouns have a two-way distance contrast for proximate and remote forms. The proximate forms are used to refer to something that is very close, often within reach, whereas the remote forms are used for things that are further away and out of reach. As for the first and second person personal pronouns, there are three case forms: nominative, accusative and ergative. However, unlike the first and second person pronouns, the singular genitive forms are not syncretic with the ergative. The different forms can occur as determiners in a noun phrase, or as pronouns - for the most part these are the same forms, but in the accusative singular different forms are used. Example 28 illustrates this difference.
The usage of the different case forms is similar to that of the first and second person personal pronouns: the nominative is used for the subject of imperfective, transitive sentences (example 29), and for the subject of all intransitive sentences (example 30). The accusative is used for the object of transitive sentences (example 31), and the ergative for the subject of perfective, transitive constructions (example 32). This results in a nominative-accusative alignment pattern for imperfective constructions, and a tripartite pattern for perfective, which will be discussed further in Section 4.8.3.

Table 8: Dative forms of demonstrative pronouns and determiners in Sauji, according to Buddrus (1967: 41-43).

<table>
<thead>
<tr>
<th>Proximate</th>
<th>Remote</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Determiner</td>
</tr>
<tr>
<td></td>
<td>Singular</td>
</tr>
<tr>
<td>DAT</td>
<td>la</td>
</tr>
<tr>
<td></td>
<td>lase</td>
</tr>
<tr>
<td>POSTP</td>
<td>lase</td>
</tr>
<tr>
<td></td>
<td>lase</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The dative forms require again some more discussion. Unfortunately, there were not enough examples in the field data to provide the complete paradigm, which is why it was necessary to rely entirely on Buddruss’ (1967) description here. The paradigm that he provided is shown in Table 8. As is has already been discussed for the nouns and the first and second person personal pronouns, it is not entirely clear if the dative forms containing te should be analyzed as dative, or rather a postpositional case form that is used together with the postposition te‘to’. In order to say more about this, however, more data would be needed which is why this matter has to be left for future studies. The postposition te‘to’ is also discussed in Section 4.6. When removing te from the pronouns, the dative forms are very similar to the postpositional forms, which, especially in the plural, are also very similar to the accusative forms presented in Table 7.

4.5 Adjectives and quantifiers

Attributive and predicative adjectives in Sauji agree with the head noun in gender and number, as examples 33 - 35 show. In the field data, the two most frequently used adjectives were ganu ‘old, big, tall’ and lawu ‘young, small’.

(33) goš-ee muškani ek nil-i drab hin-i aw rakam rakam poš-a
green-F.SG grass be.PRS-F.SG and different different flower-PL.NOM
nilbel-ee-n-ee
grow.PFV-M.PL-PRS-M.PL

‘In front of our house there is a green lawn and different flowers have grown there’ (14_HR_000519)

(34) eki dari kašan-i al-ī šeš paran-o al-o
one beard black-F.SG be.PST-F.SG head white-M.SG be.PST-M.SG

‘[One man had] a beard that was black and hair that was white.’ (18_BS_000709)

(35) sau grom-e terčapera ade law-ee aw gan-ee grom-a hin-ee
Sau village-OBL around some small-M.PL and big-M.PL village-PL.NOM be.PRS-M.PL

‘Around the Sau village there are some small and big villages.’ (20_NU_000819)

Comparative forms of adjectives are not very frequent - there are older forms such as ganeeroo ‘older’, but there appear not to be used as comparatives anymore, but rather as nouns, such as ganeere ‘the elders/leaders’. Instead, the positive form of the adjective, combined with the postposition diyaan ‘from’ is used as a comparative construction - an example is provided in 36.

(36) la gooš tase goš-a diyon gaan-o hin-o
det.prox.nom house ? house-OBL from big-M.SG be.PRS-M.SG

‘This house is bigger than that house.’ (8_MOZ_000425)

4.5.1 Cardinal numbers

Sauji has a vigesimal (base 20) numeral system, part of which is shown in Table 9. The number 21 yakaan biš, for example, consists of the number 1, combined with what could be a genitive or ablative suffix (-aan), followed by the number 20 biš. 50, dešaan dubiša, consists of the number 10, again combined with -aan, and dubiša (2 x 20). This system also goes on for numbers above 100: 110 is dešaan paanjbisa (10 + gen + 5 x 20) and 120 is Šubiša (6 x 20). The system is generally very similar to Palula - but for examples
Table 9: Cardinal numbers 1-40, 50, 60, 70, 100, 110, and 120 and their composition in Sauji

<table>
<thead>
<tr>
<th>Numeral</th>
<th>Sauji form</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>yak</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>du</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>ɬaa</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>čoor</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>paanǰ</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>što</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>saat</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>aaš</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>nu</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>deš</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>yekaaš</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>baaš</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>troyiš</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>čadiš</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>panjiš</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>ṣuṛaš</td>
<td>16</td>
</tr>
<tr>
<td>17</td>
<td>satiyaš</td>
<td>17</td>
</tr>
<tr>
<td>18</td>
<td>aštiyaš</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>uriš</td>
<td>19</td>
</tr>
<tr>
<td>20</td>
<td>biš</td>
<td>20</td>
</tr>
<tr>
<td>21</td>
<td>yakaan biš</td>
<td>1 + 20</td>
</tr>
<tr>
<td>22</td>
<td>duwaan biš</td>
<td>2 + 20</td>
</tr>
<tr>
<td>23</td>
<td>ɬaawaan biš</td>
<td>3 + 20</td>
</tr>
<tr>
<td>24</td>
<td>čuraan biš</td>
<td>4 + 20</td>
</tr>
<tr>
<td>25</td>
<td>paanǰaan biš</td>
<td>5 + 20</td>
</tr>
<tr>
<td>26</td>
<td>ṣuwaan biš</td>
<td>6 + 20</td>
</tr>
<tr>
<td>27</td>
<td>saataan biš</td>
<td>7 + 20</td>
</tr>
<tr>
<td>28</td>
<td>aašaan biš</td>
<td>8 + 20</td>
</tr>
<tr>
<td>29</td>
<td>nuwaan biš</td>
<td>9 + 20</td>
</tr>
<tr>
<td>30</td>
<td>dešaan biš</td>
<td>10 + 20</td>
</tr>
<tr>
<td>31</td>
<td>yakaashaan biš</td>
<td>11 + 20</td>
</tr>
<tr>
<td>32</td>
<td>baašaan biš</td>
<td>12 + 20</td>
</tr>
<tr>
<td>33</td>
<td>troyišaan biš</td>
<td>13 + 20</td>
</tr>
<tr>
<td>34</td>
<td>čadišaan biš</td>
<td>14 + 20</td>
</tr>
<tr>
<td>35</td>
<td>panjišaan biš</td>
<td>15 + 20</td>
</tr>
<tr>
<td>36</td>
<td>ṣurešaan biš</td>
<td>16 + 20</td>
</tr>
<tr>
<td>37</td>
<td>satiyaashaan biš</td>
<td>17 + 20</td>
</tr>
<tr>
<td>38</td>
<td>aštiyaashaan biš</td>
<td>18 + 20</td>
</tr>
<tr>
<td>39</td>
<td>urišaan biš</td>
<td>19 + 20</td>
</tr>
<tr>
<td>40</td>
<td>dubiša</td>
<td>2 x 20</td>
</tr>
<tr>
<td>50</td>
<td>dešaan dubiša</td>
<td>10 + (2 x 20)</td>
</tr>
<tr>
<td>60</td>
<td>ɬaabiša</td>
<td>3 x 20</td>
</tr>
<tr>
<td>70</td>
<td>dešaan ɬaabiša</td>
<td>10 + (3 x 20)</td>
</tr>
<tr>
<td>100</td>
<td>paanǰbiša</td>
<td>5 x 20</td>
</tr>
<tr>
<td>110</td>
<td>dešaan paanǰbiša</td>
<td>10 x (5 x 20)</td>
</tr>
<tr>
<td>120</td>
<td>ṣubiša</td>
<td>6 x 20</td>
</tr>
</tbody>
</table>
in words where the base 20 is added to a numeral, such as 21, the order is inverted in Palula, i.e. is 20 + 1 instead of 1 + 20 in Sauji. Another common pattern for numerals in the region is that especially higher numerals (e.g. above 100) are expressed with loans from languages of wider communication, such as Pashto, which is for example the case in Palula. It might be possible that this is also the case for Sauji - in one of the texts, the consultant used a Pashto loan for the number 700, as shown in example 39. Here, at least the 100-component sawa appears to be a loan from Pashto, whereas the 7-component is saat ‘7’.

As examples 37 and 38 show, the numeral is placed before the noun in a sentence. This seems to be consequent for all numerals, as far as the data indicates.

(37) gozel-e bi daš menat-a bat giri se tayar baan-u
knead.PFV-F.PL then? ten minute-PNOM after then 3SG.REM.NOM prepare become.IPFV-M.SG
‘After kneading it [the dough] for ten minutes, it is ready.’ (10_MAN_000512)

(38) čur yon-e baan-e
four month-PL become.IPFV-F.PL
‘Four months have passed.’ (17_NU_000613)

(39) la grom-e taqriban saat-sawa goš-a xalaka zahir ša aw
det.REM.NOM village-LOC approximately seven-hundred house-PL people Zahir Sha and
dawudi yarime waxta wosedal baan-al-ee
Dawud government time live? become/go.IPFV-PST-M.PL
‘At the time of Zahir Sha’s and Dawud’s government, 700 families were living in the village.’ (19_NU_000819)

4.6 Postpositions

Sauji has a number of postpositions, the most common one being tee ‘to’, which is shown in example 40. As far as the data indicates, there are no prepositions. Semantically, postpositions are mainly used to express spatial relations, but some of them are also used to express temporal, causal or instrumental relations.

Examples 40 - 42 show the usage of some postpositions. tee marks the goal of an action, such as ‘say to’ or ‘give to’. diyaan is used for the opposite: the source of an action, for example whom something is taken from. dee is an instrumental postposition.

(40) ek garazar-ee čeduan-ee tee men-oo […]
one lender-OBL borrower-OBL to say.IPFV-M.SG
‘One lender said to a borrower: […]’ (17_NU_000613)

(41) mušaar-ee dag-i teri diyaan rupa-ya mušarî til-ee
thief-OBL old-F.SG woman(OBL) from money-PL.NOM thief do.PFV-M.PL
‘The thief stole money from the old woman.’ (51_FR_170405)
(42) manus-ee para-ye deee guru jul-e si gan ți-u
man.OBL rope.OBL with horse tree.OBL together tie do.PFV-M.SG

‘The man tied the horse to the tree with a rope.’ (51_FR_170405)

Other postpositions that occurred in the field data and their approximate meanings are listed in 43.

(43) a. rate ‘on’ (spatial relation, on top of)
   b. keči ‘near’ (spatial relation, near or next to)
   c. patkeṇo ‘behind’ (spatial relation, behind)
   d. menji ‘between’ (spatial relation, between)
   e. muškani ‘in front of’ (spatial relation, in front of)
   f. sači ‘because of’ (causal relation)

Buddruss (1967: 45-46) mentions some additional postpositions such as di ‘to, towards’, gošte ‘similar to’, keeni ‘towards, in the direction of’, šaa ‘into’, si ‘together with’, muxamšuni ‘in front of’, toopee ‘under’, pa/patuno ‘after (temporal)’ and ăadee ‘in the direction of’.

Many of the postpositions in Sauji have cognates in Palula, however, rate ‘on’ (Morgenstierne 1950: 49) could also be a loan from Gawarbati.

Most of these postpositions are used with the postpositional case of the pronoun, or the oblique case of the noun - the latter has been illustrated by the previous examples. However, Buddruss (1967: 45) also reported that keči ‘near’, muxamšuni ‘in front of’, muškani ‘in front of’, and patkeṇi ‘behind’ are used with pronouns in the genitive. In the current data, patkeṇo was used with the postpositional case of the pronoun, as example 44 shows. Due to the syncretism of genitive and oblique case marking of nouns, it is difficult to tell whether the non-pronominal forms that are used with postpositions should be considered genitive or oblique. An example is provided in 45. pa and patuno occur together with the ablative, according to Buddruss (1967: 45).

(44) kučuroo […] tasi patkeṇo baan-oо aw čambaan-o
    dog 3SG.REM.GEN behind walk.PFV-M.SG and bark.PFV-M.SG

    ‘The dog walks after him [the fox] and barks.’ (16_NU_000613)

(45) aw kučuroo .chkai-u keči aw lambari kukur-u keči sot-i
    and dog goat-PL.OBL near and fox hen-PL.OBL near sleep.PFV-F.SG

    ‘The dog slept near the goats and the fox near the hens.’ (16_NU_000613)

As already mentioned, Buddruss (1967) has often considered țee a part of a dative case suffix or pronoun, but also acknowledges that it is a postposition. Looking at the examples in the field data and in Buddruss (1967) data, it is clearly possible to treat țee like any other postposition, as it behaves exactly the same way, i.e. is used together with the postpositional case of the pronoun or oblique case of the noun. This matter links of course to a much larger discussion about the grammaticalization of postpositions to case suffixes or clitics, which will not be examined any further here. Considering that țee appears to occur much more frequently than any other postposition, it is understandable why Buddruss (1967) interpreted it as a dative suffix. However, for the purpose of this thesis, it will be considered a postposition.

*except for the genitive in the singular, which sometimes takes the -an suffix instead
4.7 Verbs

In this section, the general properties of verbs in Sauji are outlined (4.7.1). Furthermore, verb morphology is discussed in Section 4.7.2 (including possible conjugation classes), and the function of the different tense and aspect categories is described in Section 4.7.3 - as far as the evidence in the field data allows. It is important to note that the labels used for the different tense and aspect categories are only preliminary, as it has not been possible to analyze the exact function of all of them due to limitations given by the data. For more clarity, these tentative labels of the different categories, are always written in cursive letters.

4.7.1 General properties

The verb always occurs clause-finally and agrees in gender and number with either the subject or the direct object of the sentence. The agreement is based on the imperfective and perfective aspect: constructions with a verb in the imperfective follow accusative alignment, i.e. in transitive sentences the verb agrees with the subject, but constructions with a verb in the perfective follow ergative alignment, i.e. the verb agrees with the object of the transitive sentence instead. This is discussed further in Section 4.8.1.

Light verb constructions consisting of a verbal component with little meaning itself, and a non-verbal component are very common in Sauji, as in many other languages in the region as well, including Palula (Liljegren 2010, 2016). Two examples are provided in (46) and (47). These constructions will not be discussed further in this thesis, but it is to assume that they function in a similar way as in e.g. Palula, which has been extensively discussed by Liljegren (2010).

(46) ek zangali lumbari aw kučuri yang bil-i
    one forest fox and dog battle/conflict(F) become.pfv-f.sg
    ‘Once in the forest there was a fox and a dog that were fighting.’ \[16\_NU\_000613\]

(47) mi jitak-u si kumaak til-i
    1sg.erg boy-obl.pl together help do.pfv-f.sg
    ‘I helped the boys.’ \[51\_FR\_170405\]

4.7.2 Verb morphology and conjugation classes

This paragraph concerns the morphological composition of the different TMA categories in Sauji - their function will be explained in Section 4.7.3.

Formation of Present Imperfective and Past Perfective forms Verbs in Sauji have a perfective and imperfective stem, which other tense and gender/number suffixes are added to in order to form the different tense and aspect categories. Based on stem formation, four conjugation classes were identified, as presented in Table 10. Verbs in the first class form the imperfective stem with the root of the verb and the suffix -aan/-een, whereas the perfective stem consists of the root and the suffix -il. The verbs in the second class form the imperfective stem with the root and the suffix -uw-aan (-uw- is a causative suffix, discussed below), and the perfective stem with the root and -al, leaving out -uw-. Verbs that belong to the third class form the perfective with the root and the suffix -t. If the root ends on a consonants, a vowel (i?) is added before the t-suffix. The imperfective stem consists of the root and -aan. Finally, there are also a number of suppletive verbs, which have completely different roots for the imperfective and perfective stem, but in most cases the suffixes that are then added to these roots are similar to the ones in other verb classes, as the example in Table 10 also shows.
Table 10: Paradigm for stem formation in Sauji verbs.

<table>
<thead>
<tr>
<th>Class</th>
<th>Example</th>
<th>Imperfective stem</th>
<th>Perfective stem</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (-il)</td>
<td>‘to do’</td>
<td>thaan- (m)/ theen- (f)</td>
<td>thil-</td>
</tr>
<tr>
<td></td>
<td>‘to kill’</td>
<td>moraan- (m)/ moreen- (f)</td>
<td>moril-</td>
</tr>
<tr>
<td>2 (-al)</td>
<td>‘to eat’</td>
<td>khuwaan- pačowaan-</td>
<td>khaal- pačoaal-</td>
</tr>
<tr>
<td></td>
<td>‘to cook’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 (-t)</td>
<td>‘to give’</td>
<td>daan- suaan- (m)/ sueen- (f)</td>
<td>dit- sot-</td>
</tr>
<tr>
<td></td>
<td>‘to sleep’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 (suppletive)</td>
<td>‘to come’</td>
<td>yaan- (m)/ yeen- (f)</td>
<td>wol- baan-</td>
</tr>
<tr>
<td></td>
<td>‘to go’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Present Imperfective and Past Perfective categories are formed by simply adding a gender/number-suffix to the imperfective or perfective stem. These markers are consistently -oo/u (masculine singular), -i (feminine singular), -ee (masculine plural), and -e (feminine plural), which we already know from the nouns.

It might be instructive to further discuss the nature of the different roots, suffixes and stems further, since it is not entirely clear what should be regarded a stem or a tense/aspect suffix, based on the available evidence. In Palula, where the verb forms are very similar, -áan is regarded a present tense suffix that is added to the imperfective stem (e.g. phed-áan-u ‘reach, arrive’), whereas the the perfective stem includes -il (e.g. phedíl-u ‘reach, arrive’) (Liljegren, 2016: 210). However, in Palula, forms such as phed-áan-u are regarded as present tense (which is based on the imperfect stem), and -áan is a present tense marker. Forms such as phedíl-u are perfective forms which do not include a tense marker (Liljegren, 2016: 253-254). This analysis is likely to work in a similar way for Sauji - however, since the function of the different tense/aspect forms in Sauji is not entirely clear (discussed in the following section), it is not possible to say if -aan should be regarded a present tense suffix, or is part of the imperfective stem. For the purpose of this thesis, -aan will be regarded as part of the imperfective stem, and the part of the verb that comes before it will be regarded the verb root, even tough another analysis might be possible as well.

Formation of other TMA categories

Except for the future, the other tense and aspect categories are based on the Present Imperfective and Past Perfective forms by adding tense suffixes to them. These tense suffixes are most likely historically related to the present and past form of ‘to be’: alo and hino (for the masculine singular) respectively. hino is shortened to -no when added as a suffix. Examples of the different forms are provided in Table 1. The Present Imperfective and Past Perfective forms are also repeated, in order to ease comparison. The previously discussed conjugation classes are only relevant for the imperfective and perfective stems - additional suffixes that are added to these stems appear to be the same, independent of the conjugation class.

The Past Imperfective is composed of the imperfective stem, as well as the past tense marker -al- and a gender/number-suffix. The Perfect consists of the perfective form (including the gender/number-suffix this time), -n-, and again a gender/number-suffix. The Pluperfect is composed by adding -al- and a gender/number-suffix to the perfective stem.

The future tense is not based on the perfective or imperfective stem. Instead it is composed of the verb root (for suppletive forms the root of the imperfective stem), -om(n)- and again a gender/number-suffix. Buddruss (1967: 53-54) argues that this form originates in an old subjunctive form, which is not used in Sauji anymore. This subjunctive form still had person agreement, but for the formation

---

7 from Buddruss (1967: 67)
8 Buddruss (1967: 53) was still able to obtain a few examples from his consultant, however, he already noted that the consultant often wanted to use the present (imperfective) form instead. In the current data, the subjunctive did not occur at all.

---

28
of the future it is eliminated by adding -n- and the usual gender/number-suffixes to the earlier first person singular subjunctive form (which ends in -m). As for the Perfect form, the -n- suffix most likely originates in the imperfective form of ‘to be’ hino. (For further discussion and comparison to Palula, see Liljegren (2016: 226-229)) What is particularly interesting here is that Buddruss (1967) already noted an assimilation (or what he called “phonetic shortening”) in some words, for example khamno ‘will eat’ was also pronounced khomno by his consultant (Buddruss, 1967: 54). In the current data, there appears to be a further shortening in e.g. yomo/yomi ‘will come’ or aṭomo ‘will bring’, where the -n- is not pronounced anymore either.

The imperative only occurs a few times in the field data, but all forms contained the suffix -ee, added to what appears to be the root of the verb. Buddruss (1967: 54-55) also mentions that the verbs of the second class add -a to the root to form the imperative (as in paša ‘show!’ or khaa ‘eat!’), or in some cases the imperative is identical with the verb root.

Valency addition For some verbs in Sauji, valency can be increased by adding the suffix -uw directly to the imperfective stem, resulting in causative forms such as pašu(w)anu ‘show’ (or ‘make see’, from pašaanu ‘see’) (Buddruss, 1967: 47,123). This difference is also visible in the perfective stem based forms, as they form the Past Perfective on -al. However, there appear to be other forms as well - the causative form of muro ‘died’ is morilo ‘killed’. It is therefore likely that there are other (regular or irregular) ways to increase (or even decrease) the number of arguments that a verb can take, also considering that there are a number of different ways to do so in the closely related Palula language (Liljegren, 2016: 235-241).

4.7.3 Tense, mood and aspect categories

In this section, the function of the different TMA categories in Sauji is described, as far as the data allows for.

Present Imperfective The Present Imperfective is used for ongoing events in the present - sometimes also with a habitual meaning. It is therefore a combination of the present tense and the imperfective aspect. Example 48 shows that it can be used to describe a state in the present. However, it appears also possible to use it for more habitual meanings, as illustrated in examples 49 and 50 show. Example 49 is
uttered in a procedural account of how the speaker and his family bake bread, and example 50 is taken from a short description of the garden of the speaker.

(48) laa anguri jaleen-i  
det.sg.prox.nom finger hurt.ifv-f.sg  
'This finger hurts.' (56_FR_170403)

(49) ma bazaar-ee pišo ataan-o  
1sg market-obl flour bring.ifv-m.sg  
'I bring flour from the bazaar.' (10_MAN_000512)

(50) aw har degar aw laalapare ma i talaan-oo  
and every evening and morning 1sg.nom water pour.ifv-m.sg  
'And every evening and morning I water it [the lawn].' (14_HR_000519)

**Past Perfective** The *Past Perfective* is used for events that have been completed prior to the time of the utterance. Again, the evidence is not entirely clear, but when comparing this form to other constructions, in particular the *Past Imperfective*, it becomes slightly more clear. The *Past Perfective* stands in direct contrast with the *Past Imperfective*, which is used for uncompleted/habitual events in the past. It is therefore likely that the *Past Perfective* is only used for completed events, i.e. is to be interpreted as a perfective aspect form. Examples 51 and 52 show the use of this form for completed events.

(51) mi teeri moril-i  
1sg woman kill.ifv-f.sg  
'I killed the woman.' (36_S_100213)

(52) aw kučuroo čhaal-u keči aw lambari kukur-u keči sot-i  
and dog goat-pl.obl near and fox hen-pl.obl near sleep.ifv-f.sg  
'The dog slept near the goats and the fox near the hens.' (16_NU_000613)

**Past Imperfective** The *Past Imperfective* functions, as previously mentioned, as a direct opposition to the *Past Perfective* form in the sense that is used for ongoing or habitual events in the past. Two examples are provided in 53 and 54.

(53) kučuroo aw lumbari sagardan geraan-al-ee  
dog and fox nervous walk.around.ifv-pst-m.pl  
'The dog and the fox were walking around nervously.' (16_NU_000613)

(54) la grom-e taqriban saatawa goš-a xalaka zahir ša aw dawudi  
det.rem.nom village approximately seven-hundred house-pl. people Zahir Sha and Dawud  
yarime waxta wosedal baan-al-ee  
government time live? become/go.ifv-pst-m.pl  
'By the time of Zahir Sha’s and Davud’s government, about 700 people were living in the village.' (19_NU_000819)
The functions of the remaining forms are slightly more difficult to describe, as they did not occur very frequently in the data (and in some cases only in the questionnaires), which has made it difficult to detect a pattern. However, an attempt to describe their functions will be made, but the analysis should be read carefully.

**Perfect** The *Perfect* is, as far as the data indicates, used for completed events in the past, that in some way are still relevant for the present. An example from the field data is provided in (55).

(55) špank-ay men-oo kūkur-ee phanti toratī šahedi man-ani aw
shepherd-obl say.PFV-M.SG chicken-GEN feather ? ? say/speak/read.PFV-F.SG and
kūkur thi khaal-o-n-oo
chicken you eat.PFV-M.SG-PRS-M.SG

‘The shepherd said: the feather is proof against you, you have eaten the (male) chicken.’ (16_NU_000613)

**Pluperfect** The *Pluperfect* only occurred in questionnaires in the field data - is is therefore difficult to say anything about its function without a context. However, looking at the examples that Buddruss provided in his texts, the *Pluperfect* appears to be used for completed events in the past, that still have relevance for another point in the past. Example (56) illustrates this - here the speaker talks about an event in the past, where the poet refers to another, earlier event in the direct speech.

(56) šair-e men-oo “ ti mafte men-aal-oо […] ”
poet-obl say.PFV-M.SG 2SG.ERG 1SG.ACC-to say.PFV-PST-M.SG

‘The poet said: “you have said to me […]” (Buddruss, 1967: 67, glosses and translation added)

**Future** The future is used for describing events that will happen in the future relative to another point in time, as well as intentions to do something. An example is provided in (57).

(57) salim mi kitaab at-oos-oo
Saleem 1SG.GEN book bring-FUT-M.SG

‘Saleem will bring my book.’ (46_FS_100213)

**Imperative** The imperative is used when directly speaking to someone and giving them a command, as in example (58).

(58) mi čed ma d-ee
1SG.GEN loan 1SG give-IMP

"Give me my loan back." (17_NU_000613)

4.8 Grammatical relations

In this section, grammatical relations, i.e. verb agreement, alignment of case marking of full noun phrases, as well as pronouns, are discussed. They all concern the core arguments in transitive and intransitive sentences: A - the subject in a transitive sentence, P - the direct object (or more patient-like argument), and S - the subject in an intransitive sentence. In Section 4.8.1, the gender/number agreement of the verb is discussed. Section 4.8.2 and 4.8.3 discuss the alignment of case marking of full noun phrases and pronouns respectively, i.e. the case marking of the different arguments and in which way
they align. These sections relate to the WALS chapters 98A and 99A (Comrie 2013a,b). Unfortunately, it has not been possible to determine the respective patterns for constructions that contain a future tense form, due to the lack of sufficient examples in the data.

4.8.1 Verb agreement

In Sauji, verbs can only agree with one argument - either the subject or the object. In imperfective-based constructions with transitive verbs, i.e. the *Imperfective* and the *Past Imperfective*, they agree with the subject, as example (59) shows. In perfective-based constructions, i.e. *Past Perfective*, *Perfect* and *Pluperfect*, they agree with the object of transitive verbs instead, as illustrated in example (60). In the imperfective, A and S are therefore treated the same, resulting in an accusative alignment. In the perfective, however, S and P are treated the same, which results in an ergative alignment.

(59)  
\[
\begin{array}{llll}
    \text{la} & \text{tusoo} & \text{piyonaan-oo} \\
    3\text{SG.PROX.NOM} & 2\text{PL.ACC} & \text{KNOW.IPFV-M.SG} \\
    \text{A} & \text{P} \\
\end{array}
\]

‘Does he know you (pl)?’

(60)  
\[
\begin{array}{llll}
    \text{asi} & \text{tasee} & \text{moril-oo} \\
    1\text{PL.ERG} & 3\text{SG.REM.OBL} & \text{KILL.IPFV-M.SG} \\
    \text{A} & \text{P} \\
\end{array}
\]

‘We killed him.’

4.8.2 Alignment of case marking of full noun phrases

The alignment of case marking of noun phrases also follows a split system which depends on perfective and imperfective aspect. This relation is illustrated in Table 12.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Alignment pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperfective</td>
<td>A = S ≠ P</td>
</tr>
<tr>
<td>Perfective</td>
<td>A ≠ S = P</td>
</tr>
</tbody>
</table>

In the imperfective, A and S are marked with the nominative case. P can be marked with the oblique, but often also remains the same as the nominative, which then results in either a nominative-accusative alignment pattern, or a neutral one. Examples (61) and (62) show objects that are marked with either the oblique or that remain in the nominative. In perfective sentences, A is always marked with the oblique, whereas P can again occur in the nominative or the oblique. S usually occurs in the nominative as well, as example (63) illustrates. Examples (63) and (64) show that P can occur both as nominative or oblique.

(61)  
\[
\begin{array}{llllllll}
    \text{la} & \text{grome} & \text{sami} & \text{xalak-a} & \text{saujibid-ee} & \text{bati} & \text{thaan-ee} \\
    \text{DET.PROX.SG} & \text{village} & \text{all} & \text{people-PL.NOM} & \text{Sauji.much-OBL} & \text{speech} & \text{DO.IPFV-M.PL} \\
    \text{A} & \text{P} \\
\end{array}
\]

‘In the village, all people speak the Sauji language.’
“And every evening and morning I water it [the lawn].”

“The man saw the (female) bear.”

“The (female) bear frightened the man.”

“The snake died.”

An exception is that in some constructions, S is marked with the oblique instead of the nominative, as in example 66. Marking S with the ergative (which in the case of Sauji is included in the case of Sauji is labelled as “oblique”), is not entirely uncommon in the region - in Kashmiri [kas], for example, there are a number of intransitive verbs that take an ergative S, ‘to laugh’ included (Rehman, 2011: 224-225).

“The little girl laughed.”

4.8.3 Alignment of case marking of pronouns

The pronouns in Sauji follow, as previously indicated, a different alignment system than the noun phrases. Again, there is a difference between the imperfective and perfective sentences. In the imperfective, A and S occur in the nominative, whereas P occurs in the accusative - resulting in a nominative-accusative alignment (see examples 67 and 68). In the perfective, however, A, S and P are all different.
A appears in the ergative, P in the accusative, and S in the nominative (see examples 69 and 70). Both systems are illustrated in Table 13. Important to note in this regard is that in the singular, the first and second person personal pronouns in the nominative and accusative are the same - which is why these systems would appear as a neutral (imperfective sentences) and ergative-absolutive system (perfective sentences) instead if one only were to look at the singular forms.

(67)  
la  tusoo  piyonaan-o  
3SG.PROX.NOM  2PL.ACC  KNOW.IPFV-M.SG  
A P  
‘Does he know you (pl)?’  (36_S_100213)

(68)  
ma  yaan-u  
1SG.NOM  come.IPFV-M.SG  
S  
‘I come.’  (46_FS_100213)

(69)  
či  door  tasee  dees-oo  
2SG.ERG  yesterday  3SG.REM.ACC  see.IPFV-M.SG  
A P  
‘Did you see him yesterday?’  (36_S_100213)

(70)  
se  wol-aal-oo  
3SG.REM.NOM  come.PFV-PST-M.SG  
S  
‘He came.’  (46_FS_100213)

4.9 Sentence modification

4.9.1 Interrogation

Sauji has a number of interrogative pronouns and adverbs such as kanee ‘where’, gašeyṭe ‘why’ and ga ‘what’, ko ‘who’, keti ‘how many’, kaaree/karee ‘when’ etc. (Buddruss, 1967: 44-45). The interrogative pronoun is usually placed before the verb, as in examples 71 and 72.

(71)  
tu  kanee  beet-oo-n-oo  
2SG.NOM  where  sit.IPFV-M.SG-PST-M.SG  
‘Where are you sitting?’  (36_S_100213)

(72)  
[…]  kučure  aw  lambari  diya  topos  thil-i  kukuri  ga  bil-o  
dog  and  fox  to?  question  do.PFV-F.SG  hen  what  become.PFV-M.SG  
‘[The shepherd] asked the dog and the fox what had happened to the hen.’  (16_NU_000613)

Polar questions can begin with the question marker aya, which also means ‘yes’, and could be a calque from Dari through Pashto varieties that are spoken in Afghanistan, considering that e.g. common Persian has a similar particle (Perry and Windfuhr, 2009: 481). Example 73 shows a polar question
in Sauji, however, in many cases this marker is also left out. Questions are then still marked by the typical question intonation where the pitch first goes down a lot in the end of the sentence, and then rises slightly again. The pitch differences between interrogative and declarative sentences are presented in Figure 4. Sauji deviates from Palula in the sense that there is no sentence-final question marker (Liljegren, 2016: 403-404).

(73) aya laa tusi kitaab hin-u
      QM DET.PROX.NOM 2SG.GEN book be.PRS-M.SG

‘Is this your book?’ (56_FR_170403)

Figure 4: Intonation difference between declarative sentence (above) and interrogative sentence (below) in Sauji. Created with Praat (Boersma and Weenink, 2020).

4.9.2 Negation

In Sauji, the sentence is negated by placing a negation marker (na/ne) before the verb, as in example (74). In light verb constructions, the negation marker is placed right before the verb, but after the noun component, as in example (73).

(74) […] ek gan-oo kukuṛ ne hin-o
      one big-M.SG hen not be.PRS-M.SG

‘[...] one big hen was gone.’ (16_NU_000613)
(75)  se  kraam na  than-u
3SG.REM.NOM  work  not  do.IPVF-M.SG

‘He is not working.’ (46_FS_100213)

The negation particle can also be merged with the verb, as in example 76. It is unclear to which extent this is also possible with other verbs with an initial vowel, as instances of that are missing in the data.

(76)  eki  šeš  rati  aw  mox-e  rati  darj  n-eel-i
one  head  on  and  face-OBL  on  beard  NEG-be.PST-F.SG

‘One [of the men] did not have any hair or beard.’ (18_BS_000709)
5 Discussion

In this chapter, some of the features that have been described in Chapter 4 are discussed in relation to areal and sub-areal tendencies in the Hindu Kush region (Section 5.2.1), the closest linguistic relatives of Sauji, such as Palula, Kalkoti, and other Shina languages (Section 5.2.2), as well as next-door linguistic neighbours, such as Gawarbat and Pashto (Section 5.2.3). Furthermore, some possible (mainly phonological) changes, that appear to have happened since Buddrus’ (1967) description, are discussed briefly (Section 5.3). However, first, a number of methodological issues and restrictions by the data are reviewed (Section 5.1).

5.1 Limitations and methodological issues

Even though there was a large amount of data available for the analysis, the data itself also set a number of limitations that have possibly influenced the results. One obvious difficulty is that, as opposed to a “classical” field study, I have not collected the data myself and have therefore not been able to talk to the Sauji consultants myself, which means that it has not been possible to clarify or confirm any of the results. Another issue is that some constructions occurred very infrequently in the data (e.g. certain tense/aspect forms), which has made it difficult to analyze certain features. In a fieldwork situation, it would have been possible to get back to the speakers and try to elicit more constructions of a certain type, or ask the speakers for clarification regarding the use and frequency of the construction. On the other hand, the data contains a variety of different texts, which presumably are still very representative of Sauji.

Balancing the use of free narratives vs. elicited materials is another important issue. The analysis is based on both, but it has been attempted to rely as much as possible on the free narratives, as questionnaires always carry the risk that the consultants stick to a certain construction and that the contact language influences the results, making the materials much less naturalistic. On the other hand, free narratives are of course much more prone to variation and mistakes, and some information might get lost due to a faster pace. In addition, free narratives might not contain the full paradigm of a certain form, whereas in questionnaires, certain forms can be targeted. In some cases, these issues have made the analysis much more difficult, which is why e.g. the analysis of pronouns had to be based on questionnaire data almost entirely. These results should therefore be taken with a grain of salt.

Another important aspect to consider is that the consultants have different linguistic backgrounds, and that this might have influenced their speech as well - especially in regards to loan words. The age of the consultants is another important factor, and in addition the data ranges over a period of 18 years, which could also be a source of smaller differences in the different data sets.

5.2 Sauji in a broader typological context

5.2.1 Areal and sub-areal influences

Sauji is in many ways a typical HKIA language, sharing many features with other IA languages inside and outside the HK region. Some of the features that are, as expected, shared with the HKIA languages, but also with IA languages in general are syntactic features such as the SOV word order or the occurrence of postpositions (Liljegren, 2017). Other features that are somewhat still typical for IA languages, but are especially prominent in HKIA languages, such as large consonant inventories or tripartite pronominal case alignment (Liljegren, 2017) can also be found in Sauji. Furthermore, features such as the occurrence of several subsets of retroflex consonants, the lack of voiced aspirated consonants, and 20-based numerals that are not particularly typical for IA languages in general, but for HKIA languages (Liljegren, 2017), also occur in Sauji.

Other patterns, that have been recognized as having more of a sub-areal distribution among the HKIA languages are also present in Sauji. These include for example that aspiration contrasts and vowel naza-
lization are less prominent in the westernmost part of the HK region (Liljegren, 2017) - Sauji is no exception here, as nasalization is not phonemic and e.g. voiced aspirated consonants are not phonemically relevant either anymore.

5.2.2 Sauji compared to its closest linguistic relatives: Palula, Kalkoti and other Shina languages

Many Sauji words are Palula-Sauji cognates, especially the more basic vocabulary (such as basic nouns and verbs, kinship terms, numerals, pronouns etc.), which clearly shows the historical connection between the two languages that has been pointed out by Liljegren (2009). The same holds for the other western Shina variety Kalkoti, although here the differences are more prominent, as Kalkoti has over the years been influenced by Gawri, a language that is spoken in close proximity (Liljegren, 2013). A number of Sauji, Palula, and Kalkoti cognates are presented in Table 14, illustrating the close relationship between the languages. Sauji is, as the examples in the table also illustrate clearly, closer related to the southern Palula dialect (Ashreti), whereas Kalkoti is linked to the northern Palula dialect (Biori) (Liljegren, 2009). Compared to the rest of the Shina languages, these three varieties form their own historical and areal cluster, as it has already been pointed out in various publications (Liljegren, 2009; Bashir, 2003; Hammarström et al., 2020; Knobloch, 2019).


<table>
<thead>
<tr>
<th>Sauji</th>
<th>Palula</th>
<th>Kalkoti</th>
</tr>
</thead>
<tbody>
<tr>
<td>'is' (was)</td>
<td>hinu (alu)</td>
<td>hinu (de)</td>
</tr>
<tr>
<td>'becomes' (became)</td>
<td>baanu (bilo)</td>
<td>bháanu (bhílu)</td>
</tr>
<tr>
<td>'does' (did)</td>
<td>thaanu (thilo)</td>
<td>tháanu (thíilu)</td>
</tr>
<tr>
<td>'goes' (went)</td>
<td>baanu (go)</td>
<td>baáanu (góom)</td>
</tr>
<tr>
<td>'gives' (gave)</td>
<td>daano (dito)</td>
<td>dáanu (dítu)</td>
</tr>
<tr>
<td>'father'</td>
<td>baabu</td>
<td>báabu</td>
</tr>
<tr>
<td>'mother'</td>
<td>yeey</td>
<td>yéei</td>
</tr>
<tr>
<td>'1.sg' (nominative)</td>
<td>ma</td>
<td>ma</td>
</tr>
<tr>
<td>'2.sg' (nominative)</td>
<td>tu</td>
<td>tu</td>
</tr>
<tr>
<td>'1.pl' (nominative)</td>
<td>be</td>
<td>be</td>
</tr>
<tr>
<td>'one'</td>
<td>yak</td>
<td>áak</td>
</tr>
<tr>
<td>'two'</td>
<td>duu</td>
<td>dúu</td>
</tr>
<tr>
<td>'three'</td>
<td>taa</td>
<td>tróo</td>
</tr>
</tbody>
</table>

Besides the similarities between Sauji and other HKIA languages in general (which have been discussed in Section 5.2.1), there are also a large number of similarities between Sauji and Palula. The case marking patterns of pronouns are almost identical - the first and second person personal pronouns in Palula also have syncretic nominative/accusative forms, as well as a syncretic ergative/genitive form (Liljegren, 2016: 126). The case alignment in Palula also follows and aspect-based split system, at least for some lexical categories (Liljegren, 2016: 297). However, Sauji and Palula also differ in certain constructions - in Sauji, definite direct objects (pronouns excluded) are marked with the oblique, but indefinite objects remain in the nominative, which is not the case in Palula (Liljegren, 2016: 106-110). This particular feature is discussed further in Section 5.2.3. For a certain numbers of verbs, S occurs in the oblique instead of the nominative, which again does not occur in Palula (Liljegren, 2016: 106-110).

The tense and aspect system of Sauji and Palula are very similar as well. The morphological composition of the different tense and aspect categories is mostly the same, i.e. there is an imperfective or perfective stem which then additional tense, gender and number markers are added to. The forms in
Sauji, however, appear to be more grammaticalized, considering that e.g. the Perfect in Palula is composed of the perfective form of the verb and the verb hino ‘is’ (Liljegren, 2016: 256-257), whereas in Sauji hino has grammaticalized to the suffix -no. Another difference between the two languages is the future tense. The Palula future tense originates from an old present form, and is the only form that still has person agreement (Liljegren, 2016: 226-228). In Sauji, this form has historically been used as a subjunctive (Buddruss, 1967: 53-54), but Buddruss already noted that it was used very infrequently by his consultant, and in the current data it did not occur at all. The future tense in Sauji corresponds to the first person singular form of this subjunctive, together with the suffix -no (which is most likely a grammaticalized form of hino ‘is’), resulting in constructions such as thomno ‘will do’. Even this form has therefore been integrated into the typical gender/number agreement system in Sauji.

The similarities between Sauji and Kalkoti mainly correspond to what has already been discussed for Palula. Kalkoti also shows the same case syncretism for the nominative and accusative first and second person personal pronouns, which as well as their ergative and genitive forms (Liljegren, 2013: 146). In Sauji, however, the latter also applies to the plural forms of the first and second person personal pronouns. The composition of the different tense and aspect categories in Kalkoti is also very similar to Sauji and Palula - also considering the different -l, -t and suppletive verb conjugation classes (Liljegren, 2013: 149).

Besides the aforementioned differences between Sauji, Palula, and Kalkoti in object marking and the future tense, Sauji also deviates from Palula and Kalkoti in its phonology. Palula, for example, has kept more “conservative” features such as voiced aspirated consonants/ consonant clusters (Liljegren, 2016: 51-60), due to its relative isolation in the mountains. Sauji, however, only has a restricted number of consonant clusters, and aspiration is only phonemic for some voiceless plosives. These changes have most likely happened due to language contact with other immediately surrounding languages such as Gawarbati, and will therefore be discussed in Section 5.2.3.

The sentence final question marker that is used in polar questions in Palula (Liljegren, 2016: 403-405), is also not present in Sauji anymore - which is again, due to areal influence (see Section 5.2.3).

Looking at Shina languages outside of the western-Shina cluster, there are still a number of recognizable similarities. The historical and geographical distance to Shina varieties such as Gilgit Shina or Kohistani Shin, is of course larger, and it has to be considered that all of these language varieties have been influenced by their respective contact languages over several generations, there is no direct contact between any of the Shina varieties anymore. A major difference is for example the fact that person agreement still is present for at least some of the verb forms in some Shina varieties (Liljegren, 2013: 149, Bailey, 1924: 26-52, Schmidt, 2004: 38-45, Liljegren, 2016: 223-225), whereas it has disappeared completely in Sauji. Other differences between the varieties have also developed due to sub-areal influences in different parts of the Hindu Kush region, as already discussed in Section 5.2.1. However, the genealogical connection between the different Shina languages is for examples clearly visible in an analysis of lexical similarity (see Knobloch, 2019: 21-22, 42-43). Another very similar feature is syncretism of the nominative and accusative forms of the first and second person personal pronouns, which is not only present in the western Shina languages, but also in other Shina varieties (Liljegren, 2013, Schmidt, 2000: 211). Otherwise, the similarities are also shared with other HKIA languages, as discussed in Section 5.2.1.

5.2.3 Influences from next-door linguistic neighbours: Gawarbati and Pashto

The Sauji lexicon contains a number of loans from Gawarbati, such as rate ‘on’, rupai ‘money’, duṅgo ‘knee’, giri ‘then, again’. (Morgenstjerne, 1950: 36,49), and Pashto, such as kumaak ‘help’, topus ‘question’, xo ‘but, however’ (discourse marker), mana ‘apple’, xawar ‘flat’ (Buddruss, 1967: 96, 104, 115). Higher numerals are also sometimes expressed with the Pashto words, even though the “original” Sauji forms can still be used as well. One difficulty with identifying loans in the current data is, however, that it is not always possible to tell if they actually are loans that have replaced the original Sauji word, *For the exact internal classification of the Shina varieties, see Chapter 2 of this thesis.*
of if the consultant just used the word because they did not know the Sauji word in this moment, or are not used to talk about certain topics in Sauji. Since Sauji is mostly used at home and in the local community, there might be topics where more loans from languages of wider communication appear.

As previously discussed, the phonology of Sauji has been influenced by its contact languages, however, changes like the absence of voiced aspirated consonants (and in general the fact that aspiration is less prominent) is most likely a sub-areal feature, rather than just an influence from e.g. Gawarbati, considering that the loss/replacement of aspiration as a phonemic contrast is very common especially in the western part of the Hindu Kush region (Liljegren, 2017: 120). On the other hand, the varying use of /f/ a /pʰ/ in Sauji, as for example in musafari/ musaphari, might be due to the immediate influence from Gawarbati, where /pʰ/ has developed into /f/ (Liljegren, 2017: 120).

A grammatical feature that has most likely developed in Sauji due to the influence of Gawarbati, is the case marking of direct objects. In Sauji and Gawarbati, the direct object is marked with either the nominative or the oblique, depending on whether it is an indefinite or definite object (see section 4.3.4 and Morgenstierne, 1950: 15), which is not the case in Palula (Liljegren, 2016: 107).

As already indicated previously, the sentence-final question marker that is typical for polar questions in Palula (Liljegren, 2016: 403-404), does not occur in Sauji. Instead, polar questions are marked by a typical intonation, and sometimes, a sentence initial question marker aya is used, which might have developed under the influence of Pashto, but is probably originally form Dari, considering that e.g. common Persian has a similar particle (Perry and Windfuhr, 2009: 481).

5.3 Changes since Buddruss’ (1967) field study

Since Georg Buddruss wrote his grammar sketch of Sauji, not many changes have happened in the language, however, some of the (mainly phonological) tendencies that he observed of an ongoing change, have become more prominent in the current data. This of course includes his observation of the disappearing of certain aspirated consonants (Buddruss, 1967: 15), which has been confirmed in the current data, where voiced aspirated consonants and aspirated affricates did not occur at all. Furthermore, it seems that his observation of a phonetic shortening/assimilation of the future verb forms (Buddruss, 1967: 54), has gone a step further in the current data.

The aorist/subjunctive form that Buddruss observed and already noted that his consultant only used it very infrequently (Buddruss, 1967: 53), did not occur at all in the current data and might have disappeared entirely in modern Sauji.

5.4 Further research

Further research could for example revisit certain aspects in the grammar of Sauji that were either not possible to answer entirely in previous research, or have not been covered at all. These could for example include a further analysis of the functions of the different tense and aspect forms. Complex sentences have also not been covered with great detail in previous studies and should therefore be analyzed further in the future. The description of Sauji phonology might also benefit from further analysis, considering that e.g. suprasegmental phonology has not been covered in the current study. But even segmental phonology might be analyzed further, considering the changes that appear to have happened since Buddruss’ (1967) description in regards to e.g. aspiration, as well as the influence of Gawarbati on the Sauji phoneme inventory. Some questions could still be answered with a further analysis of the data that has been used in this study, however, at some point, more field data would be needed to provide a more complete description of Sauji. More research on Gawarbati could also provide valuable information about patterns of language use in the region, as well as the influence that Gawarbati has had on Sauji.

Another important question to answer in the future is, however, to which extent Sauji is still spoken today and where. The language is still spoken in the Sau village, but it is unclear what has happened to the Sauji speakers that were forced to relocate to refugee camps due to the war in Afghanistan. Have they returned to the village, and to which extent do they still use Sauji?
6 Conclusions

In this thesis, a selected number of phonological and grammatical features of Sauji, a relatively underdescribed IA language mainly spoken in Sau, a village located in the Hindu Kush mountains of northeastern Afghanistan. Some of the findings have already been discussed by [Buddrus 1967](#) in a previous publication, and have been confirmed by the results of the current study. Taking new data into account, it is apparent that many of the findings constitute a completely new addition to what is known about Sauji.

What are the major phonological and grammatical features of Sauji?

A number of phonological and grammatical features of Sauji have been described in Chapter 4 of this thesis. Sauji has a moderately large consonant inventory with at least 32 consonants, and 8 vowel phonemes. Length is only contrastive for /a/ and /aː/. Two-consonant clusters are restricted to certain combinations. Nouns in Sauji have inherent masculine and feminine gender, and inflect for number and case. Personal pronouns occur in the nominative, accusative, ergative, genitive, and postpositional case, where the genitive and ergative forms of the first and second person personal pronouns are syncretic, as well as the nominative and accusative form of the first and second person singular pronouns. The third person personal pronouns correspond to the demonstrative pronouns and have a two-way distance contrast. Adjectives agree with the noun they modify in gender and number and are placed before the noun. Numerals have a vigesimal base. Sauji has a number of postpositions that are used to express a variety of spatial and temporal relations and usually occur with the postpositional form of the pronoun, or the oblique form of the noun. Verbs in Sauji have an imperfective and perfective stem, which the different tense and aspect forms are based on by adding gender/number and tense suffixes to them. Verbs agree with either the subject or the object in gender and number, depending on the aspect of the verb. The alignment of case marking in full noun phrases displays a variety of patterns, again depending on an aspect-based split. Sauji has a number of interrogative pronouns and adverbs, which are typically placed before the verb. Polar questions have a typical question intonation and sometimes start with a question marker, however, the word order remains the same as in declarative sentences. Negation is expressed with a negation particle which is placed right before the verb.

Which similarities and differences does Sauji show compared to closely related languages, such as Palula and other Shina languages?

Sauji is generally very similar to the two western Shina languages Palula and Kalkoti, due to their close historical connection, but even shows some (mainly lexical) similarities with other, more distant Shina languages. The Sauji, Palula and Kalkoti lexicon contain a large number of cognates, and even grammatically they are very similar when e.g. considering the pronoun system or morphological composition of TMA categories. However, some of the tense and aspect categories are somewhat more grammaticalized in Sauji than in Palula. The future tense works differently as well, having been integrated into the typical gender/number agreement pattern in Sauji, instead of showing person agreement as in Palula. Furthermore, the case marking of the direct object in transitive sentences is different, showing either nominative or oblique case marking, depending on definiteness, which might be attributed to the close contact with Gawarbati, which shows a similar pattern. Sauji and Palula also show a number of phonological difference, such as the absence of voiced aspirated consonants in Sauji, which might again be attributed to sub-areal influence. Considering other Shina languages outside of the cluster of western Shina, there are still a number of similarities - some of them might be attributed to a larger areal influence, but e.g. certain syncrretic pronoun forms, or the formation of the different TMA categories. However, person agreement has for example been lost entirely in Sauji, in favour of the typical gender/number agreement system, as opposed to other Shina languages.
Which similarities and differences does Sauji show compared to its closest geographical neighbors, such as Gawarbati or Pashto?

Even though Sauji is generally very similar to Palula, both lexically and grammatically, there are a number of features that are likely to have come into Sauji due to areal influences and language contact with e.g. Gawarbati or Pashto. An important aspect are of course a number of loan words. Furthermore, as previously mentioned, the phoneme system of Sauji has also been influenced by Gawarbati and other neighbouring languages. In addition, the pattern of marking definite objects in transitive sentences with the oblique, whereas indefinite objects remain in the nominative, has most likely also developed under the influence from Gawarbati, as it does not exist in Palula. Finally, Sauji does not have a sentence-final polar question particle like e.g. Palula, but uses an initial question marker instead, which has probably developed under the influence of Dari via Pashto. However, many polar questions do not have this marker, but only a typical question intonation.

Are there any recognizable changes that have happened in Sauji since Buddruss’ (1967) description?

When comparing the results of the current study to Buddruss’ (1967) results, there are some mainly phonological differences recognizable. Buddruss (1967: 15) was still able to hear a weak aspiration for some voiced consonants, but in the current data, these were not present anymore. Furthermore, the future forms that occurred in the current data were often phonetically shortened, which is something that Buddruss (1967) already noticed for some words, but now it appears to have become much more frequent.
References


A Figures and Tables

List of Tables

2 Saujiconsonantphonemeswithexamples 13
3 Transcriptionofsaujiconsonantphonemesinthisthesis 14
4 Saujivowelphonemeinventorywithexamples 14
5 Casesuffixesinsauji 17
6 Personalpronounsinsauji 20
7 Demonstrativerpronounsanddeterminersinsauji 21
8 Dativeformsofdemonstrativerpronounsanddeterminersinsauji,accordingtobuddrus 22
9 Cardinalnumbers1-40,50,60,70,100,110,and120andtheircompositioninsauji 24
10 Paradigmformstemformationinsaujiverbs 28
11 Compositionoftmacategoriesinsauji 29
12 Casealignmentinsaujifullnounphrases 32
13 Casealignmentinsaujifullnounphrases 33
14 Saujii,KalkotialtandPahulacognates.DataforPahulalandKalkotibasedonLiljegren(2013:131-134) 38
15 Dataoverview 46

List of Figures

1 LocationofthesauvillageineasternAfghanistan,wheresaujis spoken.CreatedwithArcGISonline.Linktothemap:https://arcg.is/1uP10b 4
2 ClassificationoftheShinalanguages(arealandgenealogical),basedonLiljegren(2014);Bashir(2003);Hammarströmetal.(2020) 5
4 Intonationdifferencebetweendeclarativesentence(above)andinterrogativesentence(below)insauji.CreatedwithPraat(BoersmaandWeenink,2020) 35
## Data Overview

Table 15: Data overview

<table>
<thead>
<tr>
<th>Content description</th>
<th>Consultant</th>
<th>Location</th>
<th>Date</th>
<th>Type of data (processed by...)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mainly single words, some sentences noted</td>
<td>LZ</td>
<td>Kalkatak, Chitral, Pakistan</td>
<td>15/06/99</td>
<td>Transcribed in notebook &amp; entered in Toolbox (HL)</td>
<td>1_LZ_990615</td>
</tr>
<tr>
<td>Greetings, some lexical items</td>
<td>RU</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>24/04/00</td>
<td>Transcribed in notebook &amp; entered in Toolbox (HL)</td>
<td>2_RU_000424</td>
</tr>
<tr>
<td>Simple utterances, body parts, other lexical items</td>
<td>MOZ</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>25/04/00</td>
<td>Transcribed in notebook &amp; entered in Toolbox (HL)</td>
<td>3_MOZ_000425</td>
</tr>
<tr>
<td>Text (auto-biographical) “Life story”</td>
<td>MOZ</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>25/04/00</td>
<td>Audio recording (HL), transcribed, glossed, translated (HL/NK), entered in Flex (with changes) (NK)</td>
<td>4_MOZ_000425</td>
</tr>
<tr>
<td>Clauses elicited for location and case</td>
<td>MK</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>25/04/00</td>
<td>Transcribed in notebook &amp; entered in Toolbox (HL)</td>
<td>5_MK_000425</td>
</tr>
<tr>
<td>Word list (for phonological analysis)</td>
<td>MK</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>26/04/00</td>
<td>Audio recording (HL), transcribed (HL)</td>
<td>6_MK_000426</td>
</tr>
<tr>
<td>Word list (for phonological analysis)</td>
<td>SW</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>26/04/00</td>
<td>Audio recording (HL), transcribed (HL)</td>
<td>7_SW_000426</td>
</tr>
<tr>
<td>Clauses elicited for location and case</td>
<td>SW</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>26/04/00</td>
<td>Transcribed in notebook &amp; entered in Toolbox (HL)</td>
<td>8_SW_000426</td>
</tr>
<tr>
<td>Singular-plural of nouns, gender contrasts</td>
<td>MU</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>26/04/00</td>
<td>Audio recording (AN), transcribed, glossed, translated (AN/HNL/NK), entered in Toolbox (HL), entered in Flex (with changes) (NK)</td>
<td>9_MU_000426</td>
</tr>
<tr>
<td>Text (procedural) “Making bread”</td>
<td>MAN</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>12/05/00</td>
<td>Audio recording (AN), transcribed, glossed, translated (AN/HNL/NK), entered in Toolbox (HL), entered in Flex (with changes) (NK)</td>
<td>10_MAN_000512</td>
</tr>
<tr>
<td>Text (procedural) “Making a house”</td>
<td>MAN</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>12/05/00</td>
<td>Audio recording (AN), transcribed, glossed, translated (AN/HNL/NK), entered in Toolbox (HL), entered in Flex (with changes) (NK)</td>
<td>11_MAN_000512</td>
</tr>
<tr>
<td>Word list (for phonological analysis)</td>
<td>MAN</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>12/05/00</td>
<td>Audio recording (AN), transcribed in notebook (AN), transcribed (HL)</td>
<td>12_MAN_000512</td>
</tr>
<tr>
<td>Text (hortatory) “Advice to my son”</td>
<td>MOZ</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>13/05/00</td>
<td>Audio recording (AN), transcribed, glossed, translated (AN/HNL/NK), entered in Toolbox (HL)</td>
<td>13_MOZ_000513</td>
</tr>
<tr>
<td>Mini-Text (descriptive) “Grass”</td>
<td>HR</td>
<td>?</td>
<td>19/05/00</td>
<td>Audio recording (AN), transcribed, glossed, translated (AN/HNL/NK), entered in Toolbox (HL), entered in Flex (with changes) (NK)</td>
<td>14_HR_000519</td>
</tr>
<tr>
<td>Word list (for phonological analysis)</td>
<td>HR</td>
<td>?</td>
<td>19/05/00</td>
<td>Audio recording (AN), transcribed in notebook (AN), transcribed (HL)</td>
<td>15_HR_000519</td>
</tr>
<tr>
<td>Text (fairy tale) “Dog and fox”</td>
<td>NU</td>
<td>Kalkatak Camp, Chitral, Pakistan</td>
<td>13/06/00</td>
<td>Audio recording (AN), transcribed, glossed, translated (AN/HNL/NK), entered in Toolbox (HL), entered in Flex (with changes) (NK)</td>
<td>16_NU_000613</td>
</tr>
<tr>
<td>Mini-text (joke) “King and three men”</td>
<td>NU</td>
<td>Kalkatak Camp, Chitral, Pakistan</td>
<td>13/06/00</td>
<td>Audio recording (AN), transcribed, glossed, translated (AN/HNL/NK), entered in Toolbox (HL), entered in Flex (with changes) (NK)</td>
<td>17_NU_000613</td>
</tr>
<tr>
<td>Text (traditional story) “King and three men”</td>
<td>BS</td>
<td>Sau, Kunar, Afghanistan</td>
<td>09/07/00</td>
<td>Audio recording (AN), transcribed, glossed, translated (AN/HNL/NK), entered in Toolbox &amp; translated (HL/AN), entered in Flex (with changes) (NK)</td>
<td>18_BS_000709</td>
</tr>
</tbody>
</table>

Note: HL = Henrik Liljegren, AN = Ajmal Nuristani, NU = Najib-Ullah, FR = Fazal Rahman, NL = Noa Lange, NK = Nina Knobloch
<table>
<thead>
<tr>
<th>Content description</th>
<th>Consultant</th>
<th>Location</th>
<th>Date</th>
<th>Type of data (processed by...)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text (descriptive) &quot;Sau village&quot;</td>
<td>NU</td>
<td>Kalkatak Camp, Chitral, Pakistan</td>
<td>19/08/00</td>
<td>Audio recording (AN), transcribed, glossed (AN/NK), entered in Toolbox &amp; translated (HL/AN), entered in Flex (with changes) (NK)</td>
<td>19_NU_000819</td>
</tr>
<tr>
<td><strong>Text (descriptive) &quot;Region around Sau&quot;</strong></td>
<td>NU</td>
<td>Kalkatak Camp, Chitral, Pakistan</td>
<td>19/08/00</td>
<td>Audio recording (AN), transcribed, glossed (AN/NK), entered in Toolbox &amp; translated (HL/AN), entered in Flex (with changes) (NK)</td>
<td>20_NU_000819</td>
</tr>
<tr>
<td><strong>Text (historical-narrative) &quot;Russian war&quot;</strong></td>
<td>NU</td>
<td>Kalkatak Camp, Chitral, Pakistan</td>
<td>19/08/00</td>
<td>Audio recording (AN), transcribed, glossed (AN/NK), entered in Toolbox &amp; translated (HL/AN), partially entered in Flex (with changes) (NK)</td>
<td>21_NU_000819</td>
</tr>
<tr>
<td><strong>Text (historical-narrative) &quot;Two families&quot;</strong></td>
<td>NU</td>
<td>Kalkatak Camp, Chitral, Pakistan</td>
<td>20/08/00</td>
<td>Audio recording (AN), transcribed, glossed (AN/NK), entered in Toolbox &amp; translated (HL/AN), entered in Flex (with changes) (NK)</td>
<td>22_NU_000820</td>
</tr>
<tr>
<td><strong>Text  (procedural) &quot;Oven&quot;</strong></td>
<td>MK</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>01/09/00</td>
<td>Audio recording (AN), transcribed, glossed (AN/NK), entered in Toolbox &amp; translated (HL/AN), entered in Flex (with changes) (NK)</td>
<td>23_MK_000901</td>
</tr>
<tr>
<td><strong>Text (historical-narrative) &quot;Story of one man&quot;</strong></td>
<td>MK</td>
<td>Timar Camp, Dir, Pakistan</td>
<td>02/09/00</td>
<td>Audio recording (AN), transcribed, glossed (AN/NK), entered in Toolbox &amp; translated (HL/AN), partially entered in Flex (with changes) (NK)</td>
<td>24_MK_000902</td>
</tr>
<tr>
<td>Interview notes (meta data and language situation)</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>12/02/10</td>
<td>Written in notebook (HL)</td>
<td>25_S_100212</td>
</tr>
<tr>
<td>Greetings</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>12/02/10</td>
<td>Audio recording (HL)</td>
<td>26_S_100212</td>
</tr>
<tr>
<td>Word list (40 list)</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>12/02/10</td>
<td>Audio recording (HL), transcribed (HL)</td>
<td>27_S_100212</td>
</tr>
<tr>
<td>Word list (40 list with words in sentence frames)</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>12/02/10</td>
<td>Audio recording (HL)</td>
<td>28_S_100212</td>
</tr>
<tr>
<td>Word list (40 list addition: synonym for 'drink')</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>12/02/10</td>
<td>Audio recording (HL)</td>
<td>29_S_100212</td>
</tr>
<tr>
<td>Word list (40 list addition: synonym for 'knee')</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>12/02/10</td>
<td>Audio recording (HL)</td>
<td>30_S_100212</td>
</tr>
<tr>
<td>Word list (40 list - yet another time)</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL)</td>
<td>31_S_100213</td>
</tr>
<tr>
<td>Word list (numerals)</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>12/02/10</td>
<td>Audio recording (HL)</td>
<td>32_S_100212</td>
</tr>
<tr>
<td>Word list (numerals - yet another time)</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL)</td>
<td>33_S_100213</td>
</tr>
<tr>
<td>Text (descriptive) &quot;My village&quot;</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>12/02/10</td>
<td>Audio recording (HL)</td>
<td>34_S_100212</td>
</tr>
<tr>
<td>Verb questionnaire</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>12/02/10</td>
<td>Audio recording (HL), transcribed (HL)</td>
<td>35_S_100212</td>
</tr>
<tr>
<td>Pronoun questionnaire</td>
<td>S</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL), transcribed (HL), checked, glossed, entered in Flex (NK)</td>
<td>36_S_100213</td>
</tr>
<tr>
<td>Interview notes (meta data and language situation)</td>
<td>IR</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Written in notebook (HL)</td>
<td>37_IR_100213</td>
</tr>
<tr>
<td>Word list (40 list)</td>
<td>IR</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL)</td>
<td>38_IR_100213</td>
</tr>
<tr>
<td>Word list (numerals)</td>
<td>IR</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL), transcribed (HL)</td>
<td>39_IR_100213</td>
</tr>
<tr>
<td>Word list (numerals - yet another time)</td>
<td>IR</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL)</td>
<td>40_IR_100213</td>
</tr>
<tr>
<td>Historical account of Sau (in Pashto)</td>
<td>QS</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL)</td>
<td>41_QS_100213</td>
</tr>
<tr>
<td>Content description</td>
<td>Consultant</td>
<td>Location</td>
<td>Date</td>
<td>Type of data (processed by…)</td>
<td>Reference</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>------------</td>
<td>--------------------</td>
<td>-----------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Text (historical-descriptive) &quot;About Sau&quot;</td>
<td>QS</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL)</td>
<td>42_QS_100213</td>
</tr>
<tr>
<td>Word list (40 list)</td>
<td>QS</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL)</td>
<td>43_QS_100213</td>
</tr>
<tr>
<td>Word list (numerals)</td>
<td>QS</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL), transcribed (HL)</td>
<td>44_QS_100213</td>
</tr>
<tr>
<td>Word list (long)</td>
<td>QS</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL)</td>
<td>45_QS_100213</td>
</tr>
<tr>
<td>Sentence questionnaire</td>
<td>FS</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Audio recording (HL), Partially transcribed (HL), Partially transcribed, glossed, entered in Flex (NK)</td>
<td>46_FS_100213</td>
</tr>
<tr>
<td>Notes on verb paradigms and derivation</td>
<td>FS</td>
<td>Jalalabad, Afghanistan</td>
<td>13/02/10</td>
<td>Written in notebook (HL)</td>
<td>47_FS_100213</td>
</tr>
<tr>
<td>Word list (40 list)</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>02/04/17</td>
<td>Audio recording (HL/NU), Transcribed (FR/NL), checked and entered in Flex (NK)</td>
<td>48_FR_170402</td>
</tr>
<tr>
<td>Word list (numerals)</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>03/04/17</td>
<td>Audio recording (HL/NU), Transcribed (FR/NL), checked (NK)</td>
<td>49_FR_170403</td>
</tr>
<tr>
<td>Word list (kinship)</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>04/04/17</td>
<td>Audio recording (HL/NU), Transcribed (FR/NL), checked (NK)</td>
<td>50_FR_170404</td>
</tr>
<tr>
<td>Sentence questionnaire (valency)</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>05/04/17</td>
<td>Audio recording (HL/NU), Transcribed (FR/NL), checked (NK)</td>
<td>51_FR_170405</td>
</tr>
<tr>
<td>Translated text (Northwind) - 1st time</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>04/04/17</td>
<td>Audio recording (HL/NU)</td>
<td>52_FR_170404</td>
</tr>
<tr>
<td>Translated text (Northwind) - 2nd time</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>04/04/17</td>
<td>Audio recording (HL/NU)</td>
<td>53_FR_170404</td>
</tr>
<tr>
<td>Translated text (Northwind) - 3rd time</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>04/04/17</td>
<td>Audio recording (HL/NU), Transcribed (FR/NL), checked (NK), partially glossed and translated (NK), entered in Flex (NK)</td>
<td>54_FR_170404</td>
</tr>
<tr>
<td>Text (Pear Story) - video stimulus</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>02/04/17</td>
<td>Audio and video recording (HL/NU), Transcribed (FR/NL), Translated (FR/NU), Annotated (NL)</td>
<td>55_FR_170402</td>
</tr>
<tr>
<td>Demonstrative questionnaire</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>03/04/17</td>
<td>Audio and video recording (HL/NU), Transcribed and annotated (NL), entered in flex (with changes) (NK)</td>
<td>56_FR_170403</td>
</tr>
<tr>
<td>Text (self-introduction)</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>05/04/17</td>
<td>Video recording (HL/NU)</td>
<td>57_FR_170405</td>
</tr>
<tr>
<td>Elicitation of demonstratives (classroom)</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>04/04/17</td>
<td>Video recording (HL/NU), Transcribed on white-board (FR)</td>
<td>58_FR_170404</td>
</tr>
<tr>
<td>Elicitation of gender agreement (classroom)</td>
<td>FR</td>
<td>Kabul, Afghanistan</td>
<td>06/04/17</td>
<td>Video recording (HL/NU), Transcribed on white-board and in notebook (FR), Audio recording (HL/NU)</td>
<td>59_FR_170406</td>
</tr>
</tbody>
</table>
C Sample texts

This part of the appendix includes three sample texts from the data: a short text about a garden, a descriptive text of the region around the Sau village, and a procedural text about making bread. The texts are glossed as far as possible, based on the rough translations that have been made while collecting the data, and have been completed later on using the findings from the current study.

All three recordings were made by Ajmal Nuristani, and have been roughly transcribed and translated by him and Henrik Liljegren. The first text about a garden (C.1) was provided by the consultant Habib-ur-Rahman on the 19/05/2000. Unfortunately, the location of the recording is unknown. The second text about the region around Sau (C.2) was provided by Naqib-Ullah, a refugee from Sau, at the Kalkatak Camp in Chitral, Pakistan. It was recorded on the 19/08/2000. For the purpose of this thesis, it has been shortened. The third text about making bread (C.3) was provided by Mohd Anwar, another refugee from Sau, and was recorded at the Timar Camp in Dir, Pakistan on the 12/05/2000.

C.1 Grass (mini-text, 14 HR_000519)

(77) goš-ee muškani ek nil-i drab hin-i

house-OBL in.front.of one green-F.SG grass-be.PRS-F.SG

‘In front of the house, there is a green lawn.’

(78) aw rakam rakam poš-a nilbel-ee-n-ee

and different different flower-PL.NOM grow.PFV-PL-PRS-PL

‘Different flowers grow there.’

(79) aw har degar aw laalapare ma i talaan-oo

and every evening and morning 1SG.NOM water-pour.PFV-1SG

‘Every evening and morning I water it.’

C.2 Region around Sau (shortened, 20 NU_000819)

(80) sau grom-e terčapera ade law-ee aw gan-ee grom-a hin-ee

Sau village-LOC surrounding some small-PL and big-PL village-PL.NOM be.PRS-PL

‘Around the Sau village there are some small and big villages.’

(81) leka azirgala grom šergala grom sau-gari torsan zangabošo kamtili counting Azgiral village Shergal village Sau-valley Torsan Zangabosho Kamtili

‘For example Azgiral village, Shergal village, Sau valley Torsan, Zangabosho, Kamtili.’

(82) le xalaka ađe xawar aw adi-bi mor-u rati aw adi-bi
det.PROX.NOM.PL people some flat and some-also mountain-OBL.PL on and some-also

daro wasedal baan-ee

valley live become.PFV-PL

‘Some of these people are living on mountains, some of them on plains, and some in valleys.’
These people all speak Pashto.

C.3 Making bread (10_MAN_000512)

I bring flour from the bazaar.

‘When I have brought the flour, the women here in the house sieve it.’

‘When they have sieved it, they pour water to take away the particles.’

‘When they have poured the water, then they pour flour [in it].’

‘After the flour, then they knead it with both hands.’

‘After kneading it for ten minutes, it is ready.’

‘When the bread is cooked, we eat.’