1 Introduction

1.0 Preview
This introductory chapter will give an overview of the scope and importance of the present work, provide background information to the language and the people, set out the methods and various methodological considerations, and briefly talk about the sources of inspiration for and influence on this work. We start with the statement of the aim and scope of this work along with the justification for undertaking this work in Section 1.1. Then a general background to the people and the language is provided in Section 1.2, which includes information on demography, genetic affiliation, dialectal variation, language contact, literary movements, education and literacy, and domain of language use. In Section 1.3 we discuss in detail the data and the methodology used in this work. This section will elaborate on fieldwork undertaken, types of data collected (spoken/written, child/adult), equipment and software used in collecting, organizing, and annotating data. We also dwell on corpus-based methodology and the process of creating a corpus. Section 1.4 talks about certain sources that provided inspiration for this work and whose influence is reflected in the present work. This chapter ends with an outline of the structure of the present work in Section 1.5.

1.1 Scope and importance of the work
The current work is primarily a description of the verb, verbal categories, and syntactic distribution of the verb in various clause types in Bodo. Most of the work has been devoted to the analysis of verbal forms and the distribution and function of verbal formatives. The aim of the researcher is to present an account of the Bodo verb that is comprehensive, accurate, and methodologically sound.

There are several works on the Bodo language written within the same tradition as the present one namely, functional and descriptive. Notable among them is Pramod Chandra Bhattacharya’s *A Descriptive Analysis of the Boro Language* published in 1977. However, the present work is the first attempt that is
exclusively devoted to the study of Bodo verbs, and is thus more comprehensive and complete in its coverage of the phenomena of study. The verb morphology of Bodo is quite rich and complex. The agglutinating nature of the verbal morphology combined with the rich morphology makes it particularly interesting to see how forms are concatenated and semantic/pragmatic nuances are conveyed. This richness and complexity of verbal morphosyntax is characteristic of the Bodo-Garo subgroup and Tibeto-Burman languages of Northeast India to some extent, which is what attracted the attention of the researcher. The author believes that only a comprehensive study can do true justice to the system of Bodo verbs.

Besides comprehensiveness, the current work differs from most of the other contemporary works on this area in another respect. This current work is sensitive to presenting linguistic facts about Bodo verb as accurately and objectively as possible. In this regard Bhattacharya’s work, although several decades old, is highly commendable when compared to many contemporary works. Many contemporary works suffer from two serious flaws with regard to their approach to grammatical description. First, linguistic categories are uncritically borrowed from Assamese and/or English and imposed on the grammatical descriptions of Bodo. This tendency is seen in many textbooks prepared for teaching Bodo in schools and colleges. Second, a couple of recent works have taken an approach which seems to have assumed the existence of certain pre-established grammatical categories. Such works simply enumerate grammatical categories without providing any language-specific morphosyntactic characterization. One of the many examples comes from the treatment of transitivity. It is simply assumed that all languages have the three grammatical categories of intransitive, transitive, and ditransitive verbs, even though the grammatical facts of the language may not warrant such a clear-cut distinction. As we will see in Chapter 3, the notion of transitivity in Bodo is not as straightforward as it might seem: the evidence will show that a more careful, language-specific consideration is required.

The current work differs from the existing works in yet another important way. This work provides an account of the Bodo verb based on ‘natural’ or ‘spontaneous’ data consisting of around 1.1 million words, of which two-tenths are
spoken data. This is the first ever work which is based on a database of this size. The understanding and analysis of the Bodo verb has been greatly facilitated and enriched by this database on which a more elaborate discussion follows in Section 1.3.6 below. For example, we can investigate the frequency distribution of certain constructions which are functionally similar to see if there is any difference in use. In fact, we have found constructions which are functionally very close but very different in frequency. This allows us to make some diachronic inferences involving such constructions, which otherwise would not be easily possible.

Thus, the primary goal of this work is to provide a comprehensive, accurate and methodologically rigorous description of the Bodo verb. An additional goal of this work is to establish a connection between language description and language education. We believe that there is a strong need to make language description available to the community in an appropriate form, including technical linguistic terms or metalanguage and a frame of reference to enable the language teacher to teach insightfully. This can be expected only when we make a provision to provide a pedagogically appropriate course plan to help curriculum designers in formulating suitable curricula, textbook writers in preparing textbooks/primers and teachers with methodological orientation for teaching the language insightfully. A linguistically-informed teacher will be in a better position to impart code-emphasis reading instruction (Moats, 1994) to the children; as a result, the students will be able to learn the language in a more efficient and systematic manner, especially those aspects of the language that form the core part of the grammar such as the morphosyntactic aspects of the Bodo verb.

1.2 General background to the people and the language

Bodo (or Boro) is mainly spoken by the Bodo people, one of the aboriginal tribes of Assam in the north-eastern region of India (Gait 1905: 236). They have been living along with many other tribes on the northern bank of the Brahmaputra in the foothills of Bhutan and Arunachal Pradesh since time immemorial. The people are also known as Bodo-Kachari, Plains-Kachari, Mech-Kachari or just Kacharis to outsiders. Bodo is the second largest language group in Assam, after Assamese, the language of wider communication. The term Bodo or Boro is used to mean both
the language and the people. It is pronounced as /bəro/. Following established usage, the word Bodo has been used in the present work to signify both the language and the community. Bodo is an associated official language in the state of Assam and it is also one of the 22 scheduled languages that enjoy a special constitutional status in India. Although the Assamese script was generally used for writing Bodo until the early 1970’s, after the intervention of the Central Government following the Bodo movement in the 1970’s, the Devanagari script was officially adopted for writing the language.

As far as the origin of the word Bodo is concerned, according to Brahma (1998), it may have originated from “Bod”, a place in Tibet. It is believed that they might have come down to this part from the Bod country once upon a time. Bodos belong to the Indo-Mongoloid stock racially. The Bodos are also known as “Kachari” by others, but they do not like to identify themselves by that name. They want to be distinguished as Bodosa or Bodo Phisa, meaning the sons of the Bodo race.

The largest concentration of the Bodos can be seen in the Bodoland Territorial Council (BTC) areas. The BTC, an autonomous council, came into existence in 2003. It is located in the north-western part of Assam and covers approximately 1/9th of the state. The area under the jurisdiction of BTC is called the Bodoland Territorial Area District (BTAD). There are four districts under BTC, namely, Kokrajhar, Chirang, Baksa, and Udalguri which had been carved out of seven existing districts, namely, Kokrajhar, Bongaigaon, Barpeta, Nalbari, Kamrup, Darrang, and Sonitpur. BTC has been created under the Sixth Schedule of the Constitution of India. Among the different tribes found in these areas, the Bodos form the largest group. Bodo is used as a mother tongue by the majority of the population in the said council areas. At present, Kokrajhar serves as the headquarters of the BTAD.
1.2.1 Demography

According to the Census of India 2001, the Bodo population is 1,350,478 and comprises 0.13% of the total population of India\(^1\). But, according to the key leaders of the community; their population is not less than 1.5 million at present. This population is scattered all over Assam and in some parts of the north-eastern states. Even though the majority of the Bodos is found in the BTC areas which comprise the four contiguous districts of Kokrajhar, Chirang, Baksa and Udalguri, many live in the neighbouring districts as well as in the far-flung districts of the state such as Lakhimpur, Dhemaji, Karbi Anglong, and Goalpara. Moreover, they also live in Meghalaya, West Bengal (bordering Assam), in the south-eastern parts of Nepal (Jhapa district), and in the adjoining areas of Bangladesh.

In addition to the predominantly inhabiting Bodo speaking community in the BTAD areas, many other tribes, namely, the Adivasis (a group of five linguistic communities including the tea tribes), the Rabhas, the Hajongs, the Koch-Rajbangshis, the Bengalis, the Nepalis, the Biharis, the Assamese, and the Moslems are also found to be living in the so-called homeland of the Bodos. In the past, the Bodos have been peaceful in their relation with the other ethnic tribes living in the area. But in the recent decades there have been several occasions when the community has been belligerent in their relation with other communities for reasons that are mostly demographic, socio-economic and political. In particular, as a result of recent changes in the demographic pattern of the state due to influx from adjoining Bangladesh as well as internal population shifts, the community is worried and concerned about losing their land and property to people whom they regard as outsiders.

The majority of the Bodo population follows a traditional, indigenous religion called “Bathou”. Christianity entered among the Bodos towards the end of the 19\(^{th}\) century. Some are found to be practicing other religions as well.

\(^1\)http://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement4.htm
1.2.2 Genetic affiliation and history

Bodo belongs to the Bodo-Garo branch of the Tibeto-Burman family. Bodo-Garo is a closely-knit group of about a dozen languages. The classification of the Tibeto-Burman languages of Northeast India is still an area of considerable uncertainty, but most scholars agree that the nearest relatives of Bodo-Garo are the Konyak or Northern Naga languages of Nagaland and the Jinghpaw dialect chain of northern Myanmar and Yunnan (including also the Singpho language of Assam). This grouping is often called the Sal (Burling 1983) or Bodo-Konyak-Jinghpaw (Thurgood 2003) branch. Several scholars consider Kuki-Chin, and perhaps also Karbi and Meithei, to be relatively close to this branch (Bradley 1997).

Again, genetically Bodo-Garo is distantly related to the Sino-Tibetan family. Thus, Bodo belongs to the Sino-Tibetan > Tibeto-Burman > Jingpho-Konyank-
Bodo > Konyak-Bodo-Garo > Bodo-Garo > Bodo language family\(^2\). Its sister languages are Rabha, Garo, Dimasa, and Kok-Borok (Tripuri). Tiwa and Deori are also distantly related to Bodo. Dimasa, also known as Hill-Kachari, spoken in the Dima Hasao and Karbi Anglong districts, and Kokborok (earlier known as Hill Tippera), spoken in present Tripura, are closer to Bodo than any other sister languages. In this regard, Endle (1984: i) writes:

… the Kachari language has much in common with that of Hill Tippera, so much so that the two forms of speech may perhaps be regarded as simply different dialects of what is essentially the same language.

The dialects of Meche, Mech or Mes spoken in Jalpaiguri district of West Bengal and Jhapa district of Nepal are still mutually intelligible with Bodo to a great extent even today. Linguistically speaking, the Meche variety spoken in Nepal may be regarded as a western dialect of Bodo (Kiryu 2008: 2). On the basis of suggestions made by Burling (2003), the languages that belong to the Bodo-Garo family are shown with the help of the following diagram:

\[ \text{Figure 1.2} \]

\[ \text{Relationship of the Bodo-Garo languages} \]

\[ \text{Tibeto-Burman} \]

\[ \text{(Bodo-Garo)} \]

\[ \text{Garo} \quad \text{Mech} \quad \text{Bodo} \quad \text{Kachari} \quad \text{Dimasa} \quad \text{Tiwa} \quad \text{Kokborok} \]

Historically, the Bodos were known as “Kiratas” in the days of Mahabharata. They were a ruling class and used to govern in such historic places as Dimapur, Maibong, Cachar once upon a time. They were also spread throughout the Brahmaputra Valley, which is now their homeland. Commenting on the extent and duration of Bodo domination, Gait observes,

\(^2\) Ethnologue (2005b)
The wide extent and long duration of Bodo domination is shown by the frequent occurrence of the prefix \textit{di} or \textit{ti}, the Bodo word for water, in the river names of the Brahmaputra valley and the adjoining country to the west, e.g., Dibru, Dikhu, Dihing, Dihong, Dibong, Disang, Diphnag, Dimla, etc. (Gait 1905: 6)

### 1.2.3 Dialectal divergence

Though the number of dialects cannot be confirmed as yet due to the absence of a proper dialect survey, some scholars, namely, Bhattacharya (1977: 10), have broadly divided the Bodo dialects into four main areas. These are:

(a) The north-west dialect area comprising the northern regions of erstwhile Goalpara and Kamrup districts,

(b) The south-west dialect area covering the South Goalpara, Garo Hills and a few places of South Kamrup,

(c) The north central Assam dialect area comprising the districts of erstwhile Darrang, Lakhimpur and a few places of NEFA, and

(d) The Southern Assam dialect spoken in Nowgong, erstwhile North Cachar, Mikir Hills, Cachar and adjacent areas.

However, this dialect division seems to be outdated and insufficient in the sense that it has failed to take into account the varieties spoken in other parts both within the state and outside the state. For example, the Bodo dialects spoken in West Bengal and Nepal are quite different from the north-west and the south-west dialects. Moreover, it is not clear which places or districts each of these groupings has indicated, since several new districts have been carved out of the erstwhile Goalpara, Kokrajhar, Darrang, Sonitpur, and Lakhimpur districts. These places have large Bodo concentrations. An additional factor that needs to be taken into consideration is that many people have migrated from one place to the other over the several past decades, creating a fuzzy situation regarding the Bodo dialects. Based on works done by linguists like Bhattacharya (1977: 11-12) and Kiryu (2008: 2-3) on the Bodo-Garo languages, a schematization of the Bodo dialects is indicated in Figure 1.3.
Figure 1.3
Bodo dialects as spoken in the districts of Assam, West Bengal and Nepal

Bodo dialects

Western dialects (Meche)       Eastern dialects (Bodo)³

Jalpaiguri                Jhapa dialect   WAD       CAD       SAD       EAD

dialect (Nepal) (WB⁴)

The Kokrajhar variety of the language spoken mostly in the Kokrajhar,
Bongaigaon, and Chirang districts is regarded as the standard variety. As far as the
data and analysis of the present work is concerned, we have used this variety. The
Bodo scholars use this variety in their writings, broadcasts, and academic
discourse. All the Bodo dialects are mutually intelligible.

1.2.4 Language contact and influence

Northeast India and for that matter Assam, is a multilingual state. As a result of
this situation, all the languages in this region are considerably influenced by the
neighbouring languages. Like any other language, Bodo has been exposed to the
influence of politically and economically dominant languages like Assamese,
Bengali, Hindi and English over a prolonged period. A majority of the Bodo
speakers are fluent in Assamese, the language of wider communication. Many also
know Bengali, Nepali, or Hindi and those with some amount of formal education
know English as well. Moreover, it is also not unusual for some Bodo speakers to
know another tribal language such as Rabha or Dimasa.

³Here the Eastern Bodo dialects have been sub-grouped as comprising: (i) WAD: Western Assam
Dialect consisting of the present Kokrajhar, Bongaigoan, and Chirang districts; (ii) CAD: Central
Assam dialect consisting of the present Kamrup, Baksa, and Udalguri districts; (iii) SAD: Southern
Assam dialect consisting of the present Goalpara, Karbi Anglong, and Dima Hasao districts; (iv)
EAD: Eastern Assam dialect consisting of the present Sonitpur, Lakhimpur, and Dhemaji districts.

⁴West Bengal
As a result of long contact with the speakers of the dominant languages, many words have entered into Bodo from these languages. Words such as *buzai* ‘to make (someone) understand’, *tʰogai* ‘to cheat’, *pʰorai* ‘to read’, etc. are borrowed from Assamese. Words like *saza* ‘to punish’, *kʰema* ‘to pardon’, *zorai* ‘to join’ seem to be derived from Hindi. Words borrowed from English are mostly nominal substantives – *gilas* ‘glass’, *pʰaoten* ‘fountain pen’, *berensi* ‘bench’, etc. It is a matter of coincidence that some Bodo words such as *san* ‘sun’ and *bai* ‘to buy’ are the same as in English.

1.2.5 Role of literary society and the script movement

The Bodo Sahitya Sabha (acronym BSS, a literary body that was founded on 16 November, 1952) has been instrumental for the linguistic and educational development of the Bodo people since its inception. Besides holding annual conferences and regular meetings and seminars, the body has been vocal and active in their demands for the fulfilment of the linguistic and educational needs of the Bodo people. It is to the credit of the Sabha (along with a few other Bodo organizations) for generating mass awareness among the Bodos in the field of language, literature, and education. The community has been responsive to the call of the body in the effort to advance their language and literature and to strengthen their sense of identity. Bodo is a textbook example of how the assertiveness, enthusiasm and zeal of a community can help in gaining political attention leading to the socio-political recognition of their language. Commenting on this aspect in her case study of India, Edwards (2009: 22) observes

... some 22 languages have gained regional status on the grounds that they are spoken by very large numbers of people. Political pressure also plays a part. For instance, two recently recognized regional languages – Manipuri/Meithhei and Bodo – have fewer than 2 million speakers while Bhili/Bhiladi (with 9.6 million speakers) and Santali (with 6.5 million speakers) have yet to gain official recognition.
Though the Bodos form the second largest linguistic community in Assam, they did not have a formally accepted script for their language until the 1970s. Of their own choice and will, people used the Assamese, Bengali, and Roman scripts to write the language. Meanwhile mother tongue education began in Bodo in the primary level under private initiatives from the early 1960s. The selection of an appropriate script became an urgent need at that time. At that time, the entire community was in favour of using the Roman script. The BSS had to launch a vigorous mass movement for several years during 1970s in support of adopting the Roman script. Several people lost their lives during the struggle. In the end, the then Government at the centre prevailed upon the Bodos to accept the Devanagari script instead of Roman. Since that time, Devanagari has become the officially recognized script of the Bodo people.

Commenting on the efforts of the Bodo community to establish their language, literature and culture, Suniti Kumar Chatterjee observes,

There is at present a strong movement among the Bodos of Assam Valley to establish their language as a language of literature and culture, and of instruction in schools and colleges. A modern literature is being created in Bodo, with collections of folk poetry and tales, and with literary journals. The Bengali-Assamese script is used. (Chatterjee 2007 [1951]: 46)

1.2.6 Education and literacy

Bodo is the medium of instruction from the primary level up to the secondary level in the Bodo medium schools of Assam. It is also taught as a Modern Indian Language (MIL) subject up to graduation in the colleges under Gauhati University and in some colleges under Dibrugarh University and the North Eastern Hills University (NEHU). A separate and independent department for postgraduate studies in Bodo was set up in Gauhati University in 2004. The Kokrajhar campus of Gauhati University was upgraded to a full-fledged university in 2009, named Bodoland University where there is an independent department for post-graduate studies in Bodo. The Bodo departments in both universities also offer Ph.D. degrees in Bodo. Bodo has been included as a subject in the syllabi of all India
competitive examinations conducted by the Union Public Service Commission, Assam Public Service Commission, and the University Grants Commission.

Most of the Bodo medium schools are situated in the Bodo majority areas, though a few exist outside these areas as well, in particular, in the districts of Karbi Anglong, Sonitpur, Lakhimpur, Goalpara and Kamrup (Metro). These are mostly government schools. Many Bodo children living in areas where there are no Bodo medium schools have to study either in Assamese medium or English medium schools. There are also a good number of privately managed Bodo medium schools up to secondary level in the Bodo dominated areas. It has been observed that the performance of students in the private schools is better than the students in the government schools. As a result, many parents opt for a nearby private school rather than a government run school to admit their children.

1.2.7 Domains of language use

The Bodo language is used in most domains. It is the language of the bureaucrats, officials, and administrators in the BTC areas. It is used in schools, in markets, and in social and religious occasions. The school textbooks covering the entire range of elementary and secondary school subjects are prominent among Bodo publications. One or two daily newspapers and weekly newsmagazines are published with some regularity. Monthly magazines, souvenirs, annual mouthpieces in Bodo are also published fairly regularly. The language has to its credit a large number of books of poetry, drama, short stories, novels, biographies, travelogues, children’s literature and literary criticisms. Daily radio programme (for limited durations) and television broadcasts are regular, even though the coverage of such programmes is limited.

Assamese, rather than any other language, is sometimes used at home. It is also used when two persons from different communities meet and when they do not know each other’s native varieties. Though the Bodo language is vibrant in certain areas, especially in the BTAD and other Bodo dominated areas, many people in different parts of the state have almost lost their mother tongue and shifted to
Assamese. For instance, the people living in and around the areas of Guwahati, Tezpur, and Mangaldai are of Bodo origin.

1.3 **Methodology of research**

This section presents a general discussion on the corpus-based study as a methodological tool for conducting linguistic research. In traditional linguistic analysis, we often get to see that the discussion of linguistic structures is based on data derived through intuition and introspection. While this mode has been used in certain specific situation, the overall orientation of this work has been corpus based. It is corpus-based both in terms of the data collected as well as in terms of the process of analysis itself. Thus, the Bodo data of both spoken and written genres were first collected, transcribed and entered into a computer. Next, they were systematically interlinearized and analyzed through wordlist and concordances with the help of Fieldworks Language Explorer (FLEX), a linguistic software programme developed by the Summer Institute of Linguistics International (SIL).

The following sections will help to indicate the processes that have been employed to extract the grammatical information/categories from the data by using the corpus based methodology.

1.3.1 **Corpus-based research methodology**

With the increase in interest on language research on less-described languages, there has been a corresponding emphasis on ensuring that such research is based on sound empirical foundation. In this context, a corpus-based research methodology is being preferred by many researchers nowadays. A corpus is a collection of data that is either written or spoken. To define a corpus linguistically, it is “a collection of texts and parts of texts upon which some general linguistic analysis can be conducted” (Meyer 2002: xi). The size of the corpus varies according to purpose of the work. It is understood that a corpus need to be large enough in size because any strong generalization about a linguistic structure is possible only when this generalization is based on a large size of data.
1.3.2 Fieldwork

My first experience with linguistic fieldwork was in 2004 when I was a M.A. student in the Gauhati University. In consultation with Dr. Stephen Morey, a visiting linguist from Australia who was working on the Tai languages of Assam, Prof. Jyotiprakash Tamuli, the Head of the Department of Linguistics of Gauhati University arranged for me and another classmate to gain first-hand experience in the field. Accordingly, we accompanied Dr. Morey in the process of collecting data from Aiton and Turung for his project in Silonijan and Rengmai under Golaghat district. We stayed in the area for two weeks, visited several Turung villages and were actively involved in audio and video recordings of the speech of different speakers. This was my first exposure to digital recording, both audio and video, for linguistic purposes. I also learnt the technique of data elicitation from Dr Morey. The exposure to digital recordings and the experience of real fieldwork under the guidance of an expert proved to be very useful in my own fieldwork for the purpose of Ph.D. research at a later stage.

My fieldwork was carried out in several stages, basically the first two years of my research, i.e. 2008 and 2009 were fully devoted for data collection. I had to give another year (2010) for collection of data when I decided to include children’s data into my corpus. During this period I visited several places and villages in Udalguri district and Kokrajhar district. Since my work is based on the standard variety, I chose to collect data from places where this variety is spoken and from persons whom I knew to speak the variety. The selection of villages and language resource persons (LRP) in Udalguri area was not a problem since I belong to that place and also speak the same dialect. However, I had to be careful in choosing the LRPs because of the fact that the Udalguri district, in particular, the villages near the towns are inhabited by speakers of both the standard dialect and the eastern Bodo dialect, known as Sanzari. This problem does not exist in Kokrajhar. In Udalguri district, data collection was mainly done in Maidangguri, Purani Goraibari, Dhupguri, and Chandana. In Kokrajhar district, data was collected from Kokrajhar town, Central Institute of Technology, Dotoma, Bishmuri, and Khathalmuri village.
1.3.3 Data collection

The choice and nature of data used for analysis has been an important factor in planning the present work on the Bodo Verb. Since the elements and patterns of language show distinctiveness in the way they tend to cluster around specific genres, it was important to ensure that the data represented a variety of genres or styles of the standard variety. This section deals with the process of collecting data from the spoken and written varieties of the language.

1.3.3.1 Spoken data

In recognition of the more central role of speech in the life of a community, a conscious effort was made to give due importance to spoken data in the present study. Spoken data was collected from both adults and children.

1.3.3.1.1 Adult data

During the stretch of three years of fieldwork, I recorded approximately twenty-seven hours of adult speech on tape from both male and female speakers. The recorded materials are mostly from male speakers. Although most people are unable to tell me their age, I estimate that the oldest person I recorded was around seventy five. I recorded around fifty different speakers of the language and these include many different genres of language use, namely, spontaneous speech of both male and female, folk stories, historical narratives, folk songs, conversations, process description, etc. Most of the collected material consists of folk stories, historical narratives, spontaneous speech, conversation and process description. Epic stories, narratives on cultural events and festivals, songs are least represented.

1.3.3.1.2 Children’s data

Besides collecting adult data, I also collected children’s data for my work. For this purpose, I decided to collect the data from lower primary schools of the Bodo medium. I visited several primary schools in the Udalguri and Kokrajhar districts. In Udalguri district I collected children’s data from Dhupguri Lower Primary school, Barnagaon Lower Primary school, Monpur Lower Primary school, Gerua Bagariguri Lower Primary school. In Kokrajhar district I collected data from
Chikhnajhar Bathou Ashram Phoraisali located at No. 14 Bishmuri village about 35 kilometres north of Kokrajhar town. The boys and girls from whom the data was collected ranged from Class One to Class Five (ages 5 years to 9 years). I made digital recordings of approximately twelve hours of children’s language. These covered different genres of the language, namely, stories, conversation, static as well as process descriptions, personal experience, etc. Additionally, on many occasions while collecting data from children, I also used the frog story in pictures (Mayer 1969) as a stimulus.

1.3.3.2 Written data

Written data was collected as a team during 2011-12 for the purpose of the Speech Parser Tools for Indian Languages (SPTIL-Bodo) Project. The team members, in addition to the present researcher comprised Pinky Wary, Kukil Ramchiar and Pratima Ramchiar. One of the major tasks of this project was to create a corpus of written and spoken Bodo. We collected data from different written sources, viz., newspapers, magazines, scripted movies, novels, travelogues, history, literature, etc. These include many different genres, viz., editorial, fiction, drama, history, health, sports, etc. The size of the written corpus we could create during the period is more than 1 million words approximately on which this study is based. We created the corpus from published written texts only without including any electronically published materials.

1.3.4 Equipment and recording of data

For the purpose of audio recordings of the spoken data, we used two professional recording devices –

a) Sony LS10 Linear PCM recorder with a maximum recording level of 96 kHz/24 bit

b) Tascam HD-P2, a high-definition stereo audio recorder. It records at up to 192 kHz/24-bit resolution to Compact Flash media. Along with this recorder, a high quality Beyerdynamic cardioid microphone was used to capture the voice of the speakers.
In addition to the audio recordings, we also made use of Flipcam and Sony Handycam for video recordings. These equipments were made available to me by Department of Linguistics of Gauhati University. One big advantage of these equipments is that they are digital recorders and they originally record as WAVE files (.wav format). The .wav format is a widely used platform for sound files, readable by many programmes and preferred by archives such as PARADISEC and ELAR. (Morey, 2005: 36). The recorded material will be ultimately archived at the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC, see http://paradisec.org.au). 23 sound files have already been archived with the PARADISEC so far.

1.3.5 Computerizing data

1.3.5.1 Written data

The written corpus was created by retyping the selected texts in Unicode font (Devanagari Mangal) using the Keyman keyboard layout. We saved each file in .txt format which is the compatible format for the FLEx database (1.3.7.1).

1.3.5.2 Spoken data

The spoken data was first transferred from the recording device to a computer and then transcribed. For this exercise, Audacity, a software programme for recording and editing sounds (http://audacity.sourceforge.net/) was used. In addition to provisions for recording and editing sounds, this software offers multiple options such as play/loop play, pause, stop, etc. Headphones were used to listen to the sound files in Audacity and to transcribe them one by one, using the same Unicode font and keyboard layout that were used to create the written corpus.

1.3.6 Creating the Bodo corpus

The entire data that was already transcribed and saved in .txt format were copied to the FLEx database. In all, there are 962 .txt files for written data and 101 files for spoken data in this corpus. Accordingly, two separate projects were created in FLEx programme (Section 1.3.6.1) corresponding to two genres of the language for convenience of analysis.
1.3.6.1  **Fieldworks Language Explorer (FLEX)**

Fieldworks Language Explorer provides tools for dictionary development, morphological analysis/parsing, interlinearization, etc. It facilitates the recording and analysis of linguistic data and anthropological data. Once a project is created in FLEX for the language under study, it provides a well-ordered set of fields to record data. It is a powerful programme which enables the following activities:

a) Record lexical information for developing a dictionary
b) Enter words collected at a semantic domain word elicitation workshop
c) Record vernacular texts and view a concordance of words within recorded texts
d) Create an interlinear analysis of texts
e) Analyze morphology
f) Create a grammar sketch
g) Record linguistic/anthropological field notes

1.3.6.2  **Organizing data in FLEX**

FLEX helps to organize language data systematically. The main advantage of this programme is that it can deal with large amounts of data. The 962 files containing around 1 million words from different written sources have been stored in a single database, i.e., in a single FLEX project, named *Bodo-Written*. Similarly, the 101 files containing around 100,000 words from spoken data have been stored in another database in FLEX, named *Bodo-Spoken*. Figure 1.4 shows how the text data has been organized under the middle column entitled **Texts**. It also serves the database of the corpus. The column to the right of the Texts column is the complete text of the file highlighted on the left pane. The leftmost column indicated as **Texts & Words** is the navigation pane that allows the researcher to move from one activity to the other within the project for manipulation of the corpus or analysis of the data.
1.3.6.3 Interlinearizing data

Interlinearization refers to the analysis of language data by way of morphemic break, providing morphemic/lexical gloss, information about word category, and giving a free translation of the sentence under investigation. The FLEx has tools to interlinearize language data along these lines, as illustrated in Figure (1.5). Once the information is provided for each morpheme/word or phrase, the program automatically stores it in the lexicon (see Section 1.3.6.5). Most of the data have been interlinearized using the program.
1.3.6.4 Concordance search

The FLEX program has the facility of concordance search of a particular morpheme/word. This option helps the researcher to look at the various concordances/occurrences of a particular morpheme and take an informed decision regarding that morpheme. For example, the concordance of जा /zaː/ ‘to be’ is illustrated in Figure (1.6).
1.3.6.5 Lexicon

As stated earlier in (1.3.6.3), the lexicon of the language is automatically created once we start interlinearizing the data. It can be viewed by using the Lexicon tab on top of the navigation pane (leftmost column). The lexicon shows all the entries created so far with the grammatical information such as Headword, Lexeme form, Gloss and Word Category. Figure (1.7) shows the entries in the lexicon with the word गोलोम ‘to sweat’ highlighted and its details are shown on the right pane. The Entry pane allows the user to add as much information as necessary for that particular entry in the course of analysis.
1.3.6.6 The Dictionary view

It has been mentioned that FLEX provides tools for making dictionary of a language (1.3.6.1) under study. Figure (1.8) shows the dictionary view of the language which is automatically developed by the program on the basis of information provided during analysis of data through interlinearization (1.3.6.3). The software also has the facility of exporting the dictionary data to a word program for modifications as required and for printing.
1.3.6.7 Grammar sketch of the language

On the basis of information provided during interlinearization process the FLEX programme automatically generates a sketch grammar of the language under study. This facility gives a basic idea of the language such as phonemes, different morphemes, words and their categories, etc. as shown in Figure (1.9). To explore this particular provision, we need to provide as much information as we can by way of interlinearizing the data as mentioned in (1.3.6.3).
1.4 Inspiration and influences on this work

The orientation for the present work draws on ideas received through personal communication with Scott DeLancey, Robbins Burling and Stephen Morey\(^5\) during the preparation of this work. The approach to the study of Tibeto-Burman languages such as Bodo requires a different approach than for Indo-Aryan languages such as Assamese or Hindi. In the present work, the researcher has adopted an approach that has been influenced by insights of these scholars. The present work has also been inspired by the functional typological framework of West Coast Functionalists (Givon 1984a, 1984b, 1995, 2001a, 2001b) as opposed to the generative traditions. In addition, the various workshops on Field Methods

\(^5\) Scott DeLancey is Professor of Linguistics at Oregon University, U.S.A.; Robbins Burling is Professor Emeritus of Linguistics and Anthropology at the University of Michigan, U.S.A.; Stephen Morey is Australian Research Council Future Fellow at the Centre for Research on Language Diversity, La Trobe University, Australia

1.5 Structure of the present work

An overview of the structure and content of the work is as follows. CHAPTER ONE presents the scope and importance of the present work, provides background information to the language and the people, and lays down the methods, various methodological considerations, inspirations and influences on this work. CHAPTER TWO presents a literature review on Bodo verbs. Works starting from Endle (1884) to contemporary works by local scholars are reviewed with special attention to the verb. CHAPTER THREE primarily deals with the verb root – its shape and size, borrowing, and the property of transitivity. This chapter also defines the category of verb in Bodo and provides typological information on verb in Tibeto-Burman and Bodo-Garo.

CHAPTER FOUR deals with the formation of stem by adding formatives to the root. Various types of formatives, such as causatives, versatile verbs, and adverbial suffixes are described in detail in this chapter. Moreover, the question of concatenation of verbal suffixes and the underlying organizational principle behind such concatenations are addressed in this chapter. CHAPTER FIVE deals with inflectional morphology – the kind of morphology that allows a stem to function as a predicative in a predication. This chapter describes a wide range of morphemes that indicates, time, internal organization, factuality, and speech act of verbal events. These have been treated under two broad categories - TAM and Mood. The question of whether Tense, Aspect, and Mode can be established as distinct and mutually exclusive categories is also briefly addressed.

CHAPTER SIX deals with complex verb constructions, such as auxiliary verb constructions, vector verb constructions, and conjunct verb constructions. CHAPTER SEVEN deals with clausal constructions. The primary goal of this chapter is to explore the range of verb forms found in various clause types, such as complement clauses, adverbal clauses, chained clauses, and relative clauses.
CHAPTER EIGHT deals with pedagogic considerations on language teaching in the context of the teaching of Bodo verbs in particular. This chapter talks about linguistically sensitive curriculum design and code emphasis teaching of Bodo verbs. CHAPTER NINE is a statement concerning the work, and attempts an introspective assessment of the nature of contribution of the current work. It also remarks on some of the future directions that later scholars can pursue following the line of research undertaken in the present work.