Developing a Computational Framework *for the* Verb Morphology of Great Andamanese

Dissertation submitted to
Jawaharlal Nehru University
in partial fulfilment of the requirements
for the award of the Degree of

MASTER OF PHILOSOPHY

NARAYAN KUMAR CHOUDHARY

Centre for Linguistics
School of Language, Literature & Culture Studies
Jawaharlal Nehru University
New Delhi
INDIA
2006

Date 27th July, 2006

DECLARATION BY THE CANDIDATE

This dissertation titled **"Developing a Computational Framework for the Verb Morphology of Great Andamanese"** submitted by me for the award of the degree of Master of Philosophy, is an original work and has not been submitted so far in part or in full, for any other degree or diploma of any University or Institution.

(Narayan Kumar Choudhary)
M. Phil Student
CLS/SLL&CS
JNU

Centre for Linguistics

School of Language, Literature & Culture Studies

Jawaharlal Nehru University

New Delhi-110067, India

Date: 27th July, 2006

CERTIFICATE

Certified that the dissertation titled **" Developing a Computational Framework for the Verb Morphology of Great Andamanese"** submitted by **Narayan Kumar Choudhary** to the Centre for Linguistics, School of Language, Literature and Culture Studies, Jawaharlal Nehru University, New Delhi, for the award of the degree of **Master of Philosophy**, is an original work and has not been submitted so far in part or in full, for any other degree or diploma of any University or Institution.

This may be placed before the examiners for evaluation for the award of the degree of Master of Philosophy.

(PROF. ANVITA ABBI)	(DR. GIRISH NATH JHA)	(PROF. PKS PANDEY)
SUPERVISOR	CO-SUPERVISOR	CHAIRPERSON

Gram: JAYENU, **Tel:** 26704231, **FAX:** 91-011-26165886

Dedicated
to

Those who lost themselves so that the mighty would rule for the better

To the people of Andamans

The people who never complained

Never aspired to see the pinnacle

Acknowledgement

First of all, thanks to Prof. Anvita Abbi, my guide, supervisor and director of the project titled *Vanishing Voices of Great Andamanese* (VOGA). Through, Abbi we have been initiated to the field of linguistics and to the vast arena of what is called the tribal linguistics. Abbi has been crucial for our initiation to the structure of languages also. We learnt from her how to appreciate and find the truth in the structures of the languages. Typology has been a tool to map the structures of a community's mental structure and what and how they think and analyze their world.

I thank Dr. Girish Nath Jha, my co-supervisor, guide and assistant professor of Computational Linguistics. He is the person who has initiated several people to the field of NLP and AI. Without him, this undertaking could never have completed. It was he who showed and provided most of the programming part of this undertaking, I being the person who offered the algorithm and the structure of the language. It is he who initiated me to the area of machine language and making the machine understand the human language.

I am grateful to all my teachers at the Centre for Linguistics, School of Language, Jawaharlal Nehru University from whom I learnt the basics of linguistics. I am thankful to Prof. P. K. S. Pandey who taught me phonetics and phonology and guided into the area of research by showing us the valid methodology a researcher should depend upon for its sanctity to be maintained. Thanks to Dr. Ayesha Kidwai for initiating us to the field of generative linguistics- especially morphology, syntax and semantics. Thanks are due to Prof. Vaishna Narang also who initiated us to the area of applied linguistics. That is where I learnt that linguistics is not just about theories of languages, but also it can be of use to the people at large. I thank Prof. Franson Manjali because I snatched a bit of the philosophical attitude to the world from him and he did not say anything.

This project was funded by School of Oriental and African Studies, University of London under the Hans Rausing Endangered Languages Documentation Programme. Thanks are due to the funding agency. Working in this project as a research assistant under the kind and experienced guidance of Prof. Anvita Abbi has been an asset to me that will be cherished.

I should thank my team members in the project, Dr. Alok Das and Abhishek Avtans, who accompanied me in my field trip to the wilderness of Andamans. The trip to the islands and its accompanying experiences will long be in our memories.

At last, I thank all my informants, especially Nao Jr. He was the one who gave enough of the data from his past memory. Despite several complaints of loss of memory, he managed to press his past to come alive.

I thank Jagdev, my neighbour and hostel mate for a fine proof-reading of the non-technical texts in this dissertation in the last minutes.

Hey Mallu, you are never forgotten. You asked me to acknowledge you for providing the fillip to the area of computational linguistics. Well, you see here you are. Delivered.

Contents

Abbreviations Used

1	First Person
2	Second Person
3	Third Person
ABL	Ablative
ACC	Accusative
ALPAC	Automatic Language Processing Advisory Committee
APPL	Applicative
ASP	Active Server Pages
AUX	Auxiliary
BENEF	Benefactive
CAUS	Causative
C-DAC	Centre for Development in Advance Computing
CL	Clitic
CLSM	Class Marker Consonant or Thematic Consonant
COLING	COmputational LINGuistics
COND	Conditional
DEM	Demonstrative
DIREC	Directional
DIST	Distal
EGL	Gloss in English
EXCL	Exclusive
EXIST	Existential
EPV	Epenthetic Vowel
GA	Great Andamanese
GAVA	Great Andamanese Verb Analyzer
GEN	Genitive
GL	Gloss
HABIT	Habitual
IIIT	International Institute of Information Technology
IIT	Indian Institute of Tecnology
IMP	Imperative
INCL	Inclusive
IND	Indicative
INSTR	Instrumental
JSP	Java Server Pages
LOC	Locative
MB	Morphemic Break
NCST	National Centre for Software Technology
NEG	Negative
NLP	Natural Language Processing
NPST	Non-Past

OBJ	Object
PCPL	Participle
PL	Plural
POS	Part Of Speech
POSS	Possessive
PROH.NEG	Prohibitive Negative
PROX	Proximate
PST	Past
PURP	Purposive
REFL	Reflexive
REM	Remarks
SG	Singular
SOC	Sociative
STAT	Stative
STT	Speech To Text
SUBJ	Subject
TDIL	Technology Development for Indian Language
TIFR	Tata Institute of Fundamental Research
TRV	Transitivizer
TTS	Text To Speech
VD	Vowel Delition
VH	Vowel Harmony
VI	Vowel Insertion
VR	Verb Root

List of Tables, Figures and Maps

Figures

Tables

Maps

Screen Shots

Images

Chapter 1

Introduction

1.1. The Andaman and Nicobar Islands

The Andaman and Nicobar Islands and the sea in which this group of islands are situated have for long been called as *Kalapani* or Black Waters by the common mass in India as it is mentioned in the memoirs of the colonial British rule when convicts and political leaders were sent to the Cellular Jail. The punishment of *Kalapani* was considered more dreadful than the capital punishment. For it was considered that those who were sent to this place will never be able to come back. And indeed it was the case in most of the instances. Almost all the convicts sentenced for *Kalapani* stayed back there after their terms in the jail was over. They were given land by the colonial settlement there and they settled there along with the other settlers brought from other places by the British. The original inhabitants of the islands though did their best to drive these settlers out; they were always overpowered and pushed back to jungles. Eventually, by the time the British left and handed over these islands to independent India, these islands were populated with several alien communities. The major among them were the Moplahs from Malabar in Kerala, the Bhatus or Bhantus from Uttar Pradesh, Karens brought from Burma, the tribal brought from Jharkhand, and the freedom fighters along with convicts from all parts of India.

After Indian independence, this group of Islands was given the status of a Union Territory, and the area was divided in two districts namely Andaman and Nicobar. There followed a spate of systematic settlement of Bengalis from East Pakistan (now Bangla Desh). At present the Bengali community comprises the most of the total population of more than 350 thousand. Besides, there has been a surge of Tamil and Telugu opportunity seekers in these islands. However, one can find all the major Indian communities earning a living here. And, yes, this place is also a hot tourist

spot for several countries. So, one should not be surprised to see tourists from Europe spending months in the hotels or camps.

1.1.1. The Land of Andaman and Nicobar Islands

1.1.1.1. Geographical Location

Situated between 92nd and 94th meridian of East Latitude and 6th and 19th parallels of North Latitude, the range of islands form an irregular chain that seems to continue the Himalayan ranges through Myanmar's Arakan Yoma southwards in an arch over 1100 km. of sea into Sumatra.

This territory comprises islands some of which are large such as North Andaman, Middle Andaman, South Andaman, Baratang, and Little Andaman in the Andaman group and Car Nicobar in the Nicobar group. Situated in the Bay of Bengal, Andaman and Nicobar islands constitute one of the most important union territories. Geologically the islands appear to have been part of the land mass of South-East Asia comprising North-East India, Burma, Thailand, Malaysia and Indonesia.

1.1.1.2. The Terrain

Separated by numerous straits, this area contains 204 islands- 572, if all the islets and rocks be faithfully counted. The terrain is generally mountainous with long ranges of hills enclosing narrow valleys. The configuration of the land clearly points to these islands being the visible ridges and summits of sunken ranges of mountains; Saddle Peak in North Andaman at a height of 732 metres above sea level is the highest point in these islands. There are no great elevations and the slopes are moderate to steep and rugged. They are susceptible to heavy soil erosion. Flat lands are comparatively scarce. The villages in Betapur and Diglipur constitute the main flat lands in the Andaman group.

1.1.1.2.1. The Nicobar Group of Islands

The Nicobars are separated from the Andaman Islands to the north by the 150 km wide Ten Degree Channel and are 189 km from the Indonesian island of Sumatra to the southeast. The Andaman and Nicobar islands separate the Bay of Bengal from the Andaman Sea. Indira Point, south of Great Nicobar, is the southernmost point in India.

The Nicobar Islands include 22 islands of various sizes, the largest being Great Nicobar. The total land area of the chain is 1841 km². The highest point on the Nicobars is Mount Thullier at 642 m.

The population of the islands was 42,026 in 2001, roughly 65% of whom are indigenous peoples (the Nicobarese and Shompen peoples, listed among the Scheduled Tribes of India), and 35% migrants from India and Sri Lanka.

The islands in the Nicobar group are surrounded by coral reef and shallow seas. Long narrow stretches of sandy beaches are a salient feature of the topography here. Car Nicobar and Katchal are almost flat while the others have hilly terrain. In Little Nicobar and Great Nicobar in the Nicobar group, the land surface is very irregular, cut up by steep hills and valleys.

1.1.2. Andaman in Historical Records

The Andaman Islands have been by and large an isolated place inhabited by the people of Negrito stock for thousands of years. The earliest mention of these islands in any modern historical records is that of Claudius Ptolemy, Greco-Roman Geographer in 2nd Century A.D., who first published an atlas of the world and marked these islands as ‘Angdaman’. I-Tsing, a Chinese Buddhist monk, also mentioned of these islands in 672 A.D. There were also two Arabian travellers who passed through these islands. Their account makes a clear mention of these islands and its inhabitants. Marco Polo, acclaimed Venetian merchant in the 14th century, also wrote about these people, rather with a sense of fright which Portman (1899) later said was based mostly on the hearsays from the Malaysian pirates that “the men of this Island

of Angamanain have heads like dogs, and teeth and eyes likewise; in fact, in the face they are like big mastiff dogs! They have a quantity of spices; but they are most cruel generation, and eat everybody that they catch, if not their own race. They live on flesh and rice and milk, and have fruits different from any of ours."(Yule, 1903:309).

There were several other contacts made even after that namely those of Friar Odoric in 1322, Nicolo Conte in 1430, Master Cesare Frederici in1566 and several others, who described the land and the people in their own way. But all these visits were very short in description and none of them were reliable until the British came in these islands in 1788 and established a penal settlement. The British explored many parts of these islands and finally settled there with the help of labourers and convicts brought from India. Portman's 1899 work is the first comprehensive work on the land and the people.

1.1.3. The People and the Language

As said earlier, the main population there is the settlers, now addressed as Locals in the administrative reference, brought from different parts of the surrounding areas. They have also been now segmented under two categories namely the Pre-Fourty Two and the Post-Fourty Two.

The indigenous people of Andaman Islands are the Negrito people of 13 primitive tribes. Out of these, ten tribes, who were the most thriving people of the land and had linguistic affinity with each other, have now become extinct. They are now termed as 'the Great Andamanese'. However, much work has to be done on the relationship of the Great Andamanese with the other Negrito tribes of the Andaman district.
At present the division of the indigenous people residing in the Andaman district is done on a four tribe basis and are as follow. All these tribes are considered as primitive tribes.

i) The Great Andamanese

ii) The Jarawa

iii) The Onges

iv) The Sentenelese

1.1.3.1. The People of Andaman District

1.1.3.1.1. The Great Andamanese

The Great Andamanese were the most thriving people in the islands and the first to come in contact with the settlers there. There were ten tribes in this group spread throughout the Andaman group of Islands. Each of the tribe had a distinct dialect of its own. Eventually, they bore the brunt of the occupation of their land and have finally succumbed to it. These were only 23 in 1970 when they were finally settled in Strait Island, a tiny island of about 1.2 kms^2. Their number has increased now up to 50, but this is a mix population. There has been rapid shift in language and the younger generation is unwilling and non-enthusiastic to learn and use their language. There are hardly six to seven people who can speak something in their own language. Even that comes with great difficulty. Next generation transfer of the indigenous language has stopped completely. The settlement area of Strait Island has been declared as primitive tribal zone and it is not allowed to have contact with them without permission from the Adim Janjati Vikas Samiti i.e. Primitive Tribe Development Committee and the Administration.

1.1.3.1.2. The Jarawas

Till recently, they were the dreaded tribe among the settler community there. But with continuous efforts from the administration and various other people, these people have now become friendly. There are a total of 250 Jarawas living in three groups in three parts of the Andaman district, namely Middle Strait, Baratang and Kadamtala.

All these three parts are separated from each other through straits which they cross now with the help of the canoe procured as and when required by the Tribal Welfare department. They stay mainly inside the jungles. With the passage of time, the behavioral patterns of Jarawas have changed. Till the beginning of 1998, they remained hostile, but now they are coming out of the jungle quite often and are becoming friendlier. For nearly a year there is no incidence of killing of villagers by the Jarawas. On the other hand, the Jarawas are coming out from their habitat to mix with the local people. After giving them gift items like bananas, coconuts, etc., they are being sent out to live in their own natural habitat, with a view not to force them to have a taste of the civilized world. Isolated so long, the Jarawas otherwise appear to be healthy, with smooth skin, deep curly hair, long and sturdy hands and legs and sturdy bones. They are physically fit for hunting, fishing. Unlike, other tribes mentioned earlier, the Jarawas are not welfare dependent people. As nomadic tribes subsisting on hunting, fishing and gathering activities, their traditional food articles consist of boar (wild boar), turtles and their eggs, crabs and other shore animals, etc. wild pig fruits and honey.

1.1.3.1.3. The Onges

The Onges are the people residing in the two pockets of Dugong Creek and South bay in the Little Andaman, approximately 100kms south of Port Blair in the sea. It does not seem probable that the Onges may have had any kind of contact with the other tribes mentioned here. They are also diminishing in number. They are the semi-nomadic tribes and fully dependent on the food provided by nature. Like the Great Andamanese, they have also been provided with several amenities by the Administration since they became friendly. But they are also now experiencing the effects of the outsiders. They eat turtle, fish, roots and jack fruits etc. They have developed arts and crafts. The Onges can make canoes. But this tribe, presently around 100 in number, is facing the danger of extinction.

1.1.3.1.4. The Sentinelese

The Sentinelese are still a hostile tribe and does not entertain anybody on their small island of around 60km^2. They live on the North Sentinel Island and never leave this place. All attempts to make contact with these people have been futile. They are considered as an off-shoot of the Onge-Jarawa tribes which have acquired a different identity due to their habitation in an isolated place and have lost contact with the main tribes.

1.1.3.2. The People of Nicobar District

The Nicobarese, the largest group, seem to be of mixed Burmese, Malay, Mon and Shan origin. They are a friendly and cheerful lot, who do not accept money and prefer the barter system. Communing with the dead is one of the many intriguing rituals practiced in Nicobar

The Nicobar Islands are believed to have been inhabited for thousands of years. Six indigenous Nicobarese languages are spoken on the islands, which are part of the Mon-Khmer branch of the Austro-Asiatic language family, which includes Mon, Khmer and Vietnamese languages of Southeast Asia, and the Munda languages of India (Abbi, 2001). An indigenous tribe living at the southern tip of Great Nicobar called the Shompen may be of East -Asian origin.

1.1.3.2.1. The Shompens

The Shompens are another primitive tribe living in Great Nicobar. Like the Nicobarese, they belong to the Mongoloid race. The Shompens have two divisions, the smaller division being known as Mawa Shompens. They inhabit areas very close to the coastal region along the river valleys. They are very shy. They are quite intimate with the Nicobarese and of the major group of Shompens, The “hostile” Shompens are living in Alexendra and Galathia river areas and also on the east coast of the area in the interior of the Island. In the past, frequent attacks are believed to

have been made on the Mawa Shompens by the hostile Shompens. But now, such hostility has been stopped. It is probably because they have been largely reduced in number due to various diseases. The Mawa Shompens are the victims of diseases and physically very weak. With the establishment of the settlement at Campbell Bay in Great Nicobar, Shompens have been visiting the settlers and they are gradually shaking off their shyness and indifferent attitude towards the civilized people. There are several stories about their origin. Some people say they have come from the neighbouring countries in Malaysia and Indonesia.

1.1.4. Where do the Andamanese Belong?

The aboriginal people of Andaman Islands are of Negrito stock. They are also likened to people of the same stock found in the Malay Peninsula and the Philippines, namely the Aeta and Semang respectively (Pandit, 1976). There are indications of having similarities with the Veddas of Sri Lanka and the Panyans of Kerala in India (Endicott et. al. 2003). There are also indications of their genetic similarities with the Negrito people of other areas of the globe.

However, no clear results have yet come up as to since when they have been isolated in the Islands. Endicott et al (2003) also notes that "All lines of evidence—social, cultural, historical, archaeological, linguistic, phenotypic, and genetic—support the conclusion that the Andaman Islanders have been isolated for a substantial period of time."

About the linguistic affinity among the people of these Islands, scholars say that the family stands out and has no relationship with any other language family (Brown 1948, Nigam 1972). Brown also points out that even if it is taken that Andaman Islanders have closer linguistic affinities with the Aetas and Semangs of the Malaya and Philippines, they have changed and modified by the surrounding alien culture.

At present, it is commonly accepted that Andamanese forms a separate family of language, at least geographically. The Andamanese family of language is further

grouped into two- Great Andamanese and Little Andamanese (Manoharan, 1989). In the former group all the ten languages are put together, almost all of which are extinct today, except for at least one (Jero) which may be deteriorated due to disuse and some possible mixture with other dialects. All the three groups of Onge, Jarawa and Sentinelese are put into the Little Andamanese group. According to Abbi (2003) Andamanese family is divided into Western and Eastern Andamanese. The ten languages of the Great Andamanese group, namely Kede, Kol, Juwai, Pucikwar, Bea, Bale, Jeru, Bo, Sare and Khora, belong to the Eastern group. The Western group is further divided in Central and Southern. Jarawa belongs to the Central Western while the Onge and the Sentinelese belong to the Southern Western.

While Onge and to some extent Jarawa, have been studied linguistically, any kind of study on the Sentinelese is yet to come as the Sentinelese aggressively protest all kinds of encroachment on their land.

The internal linguistic affinity between the two subgroups of the Islands was attempted by Greenberg (1971:810) in his *The Indo-Pacific Hypothesis* where he suggested Andamanese links to the non-Austronesian languages of Melanesia and to Tasmanian on the basis of 35 cognates. That the two sub-groups share great morphological resemblances was also pointed out in Radcliffe-Brown (1914) and later validated by Manoharan (1983:86).

Abbi (2006) discusses in detail on the linguistic affinity among the Andaman Islanders. Based on the typological comparisons of several features of all the three groups of languages, the result states that “the overall typology of these languages appears similar”. Abbi, (2006 pp.93) Further, undertaking the analysis of Great Andamanese Abbi finds the complex verb morphology of Great Andamanese most distinguishing from Jarawa and Onge. If Manoharan considered the Andamanese the fifth language family of the Indian subcontinent, Great Andamanese “is either a distant relation of this family or may even constitute the sixth language family of the Indian subcontinent” (Abbi, 2006 pp.96).

1.2. The Great Andamanese

As said earlier, the Great Andamanese were the most thriving people of the Islands, far more in number and occupying most of the landmass in the Andaman Islands. Their estimated population in 1789 was 10,000. They are the first to bear the brunt of modern colonization starting in late 1700s. By 1901 their population diminished to 625 which further reduced to 19 in 1969. This drastic and almost immediate death blow to the primitive community due to colonization of their land and insensitive administrative treatment to bring them under control made them go unnoticed down in the history until Indian administration finally awoke, though very late, in 1970 and resettled the rest of these people (only 19) scattered at different places to a tiny island called Strait Island. Although the numbers grew up after the resettlement so much so that at present (i.e. March, 2006) their number can be shown up to near 50, most of the younger generation is the result of cross-breeding.

1.2.1. The People

The Great Andamanese people are different from the Little Andamanese group. The Great Andamanese were divided in 10 tribes, each having a distinct name and a dialect. Usually the Great Andamanese are divided in three groups according to their geographical spread: Northern, Southern and the Middle Group. Man (1932, as quoted in Manoharan (1989) gives the following sub grouping:

i)	South Andaman:	Bea, Balawa
ii)	Middle Andaman:	Puchikwar, Juwoi, Kede & Kol
iii)	North Andaman:	Jeru & Cari

While Man did not include in this list the two dialects of Bo and Khora of the North Andaman, Brown (1948), drawing two sub-groups, gives the following list of all the ten dialects:

North: Cari, Kora, Bo & Jeru

South: Kede, Kol, Juwoi, Puchikwar, Bale & Bea

Manoharan also provides a greater detail on the sub-grouping of these people as follows:

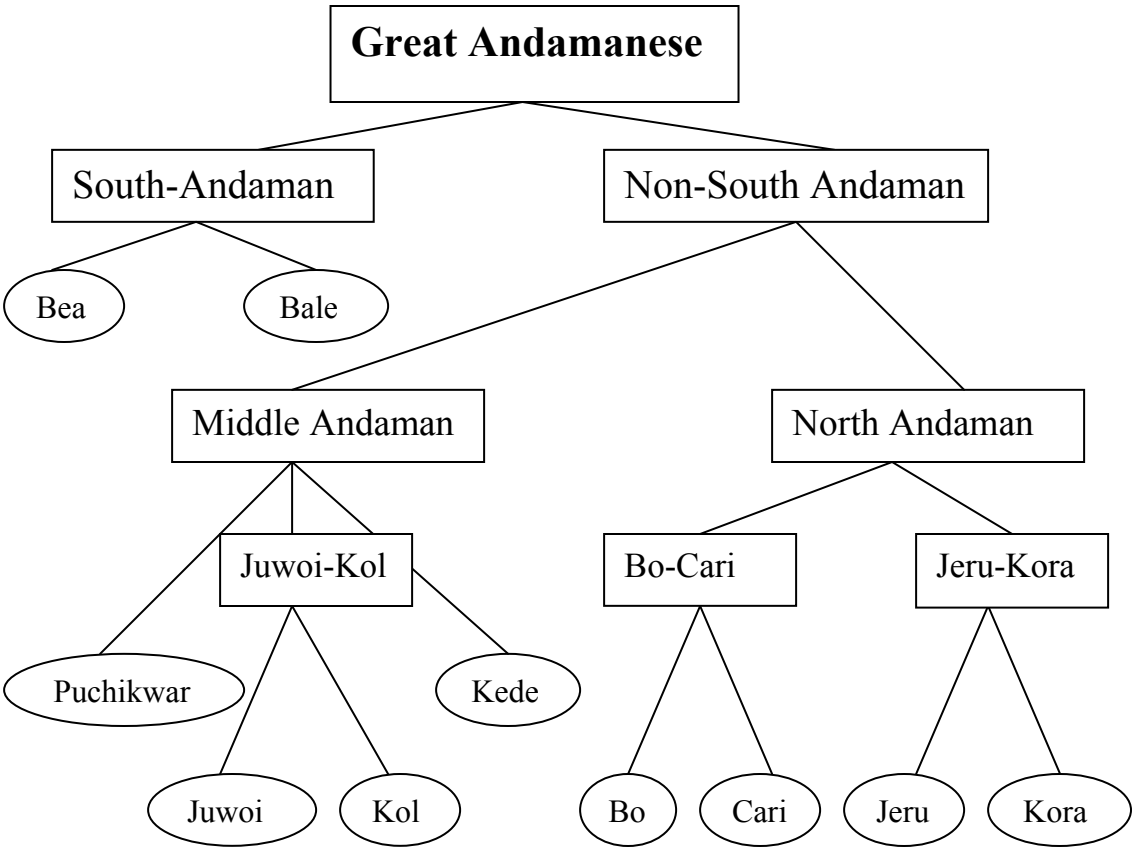


Figure1: A Stammbaum of the Great Language family

Source: Adapted from Manoharan, 1989 pp. 173

As the Great Andamanese is now on the verge of total extinction, it will be worthwhile to have a look at their demography since the colonization of the land started. The following table has been taken from the Andaman Association's website:

Tribe	**1858**	**1883**	**1901**	**1911**	**1921**	**1931**	**1951**	**1961**	**1971**	**1981**	**1995**
Aka-Cari	100		39	36	17	9					
Aka-Kora	500		96	71	48	24					
Aka-Bo	200		48	62	16	6					
Aka-Jeru	700		218	180	101	46					
Aka-Kede	500		59	34	6	2					
Aka-Kol	100		11	2	0	0					
Oko-Juwoi	300		48	9	5	0					
A-Pucikwar	300		50	36	9	1					
Akar-Bale	300		19	15	4	2					
Aka-Bea	500		37	10	1	0					
Sub-Total	**3500**	**2000**	**625**	**455**	**207**	**90**	**33**	**19**	**24**	**24**	**37**

Table 1: The Demographic Table of the Great Andamanese

Source:<Andaman.org./BOOK/>chapter seven (viewed on 22nd March, 2006) [1]

Note: the present Great Andamanese population is about 50 as in March, 2006.

The table above makes it clear along with the comments made above that the tribe is on the verge of total extinction. My five months long stay in the Andamans with the remaining of the tribe proved the fact even more emphatically. As a linguist, and as I was assigned the task of eliciting linguistic data to write grammar , I was always in want of talking to the tribe's people in their own tongue. Out of the 50 odd people that they are at present, only five or so could speak their mother tongue. Even these

[1] All the names the tribes of the Great Andamanese are prefixed with Aka-, Akar-, A- or Oko-. The usual interpretation given for this prefixing is that it means 'from the mouth or tongue' (see the web pages of Wikepedia, George Weber's Andaman.org etc). But this interpretation seems to be a simplistic one. /aka-/ is not prefixed only to the names of the languages, this comes before several other nouns such as /aka -may/ "my father" or /aka-mimi/ "his mother". This prefix also comes before verbs where it signifies third person singular object or subject clitic. For more on the clitcs and verbal prefixes, see chapter three.

few speakers are not fluent in their language. All the youngsters have started speaking the lingua franca of the area i.e. Hindi. Even the elders confer much less in their own tongue. The social tie-up amongst the tribesmen is also reducing. They have adopted all the ways and manners of the modern, metropolitan culture the area can offer.

Despite all these, they are very friendly to the outsiders. The comments made on these people earlier by the visitors to these Islands and its people that these people were cannibals or treat the outsiders with great inhumanity can only be the natural result of human curiosity and suspiciousness that any race may have for an alien race and those comments may well be treated as only misinterpretations of their hostile attitude towards any kind of alien encroachment of their land. Barring their indifference to my commitment to their language and my work, my experience with the Andamanese people of Strait Island was pleasant. I was only peeved at the indifference they showed to their own works and carelessness. This, though, one can say, is the result of their own understanding of the modern world which they know can accept them only if they accept their way of life. Resistance can no longer be one of the options.

1.2.2. The Language

When the remaining of the Great Andamanese people was resettled at Strait Island in 1970, they were only 19 in number. The majority among these were the Northern Jeru speakers. Perhaps there were one or two speakers from each of Cari, Khora and Pucikwar dialects also. The majority being Jeru, the community adopted this language as their language in common. Their present language is Jeru in which one can find words from Pucikwar, Cari and Khora mixed intricately. Linguistic data on Andamanese with any details before Manoharan (1989) is scanty. The earliest published work of importance on Andamanese language is that of Portman's *Andamanese Manual* (1887) where he has taken representative languages from all the three geographical subgroups namely Aka Bia-da or Aka Bea (North Andaman), Aka Bojigiab or Pucikwar and Aka Kede or Kede (both Middle Andaman) and Chariar (North Andaman). E.H. Man and R.C. Temple had published the *Lord's Prayer in*

South Andaman in 1877 followed by his *Dictionary of the South Andaman Language* in 1919. Radcliffe-Brown was the first to work on the North Andaman tribes whose works were published in 1914 and 1948. The next work on the Great Andamanese was attempted by Manoharan, an anthropology research scholar at Calcutta, in 1981 (published in 1989). As all of these works were done by non-linguists it is not surprising we have not yet got any comprehensive grammar of any of these languages. Manoharan's work attempts to analyze the language but fails at several points to give a proper explanation for the data elicited. Besides, there is also a lack of linguistic evidences for the rules drawn in his descriptive and comparative study.

The last study done on the languages of Andaman Islands was that of Abbi's pilot survey in 2001 which came in publication in 2003 and there are two more in the offing. This was an all linguists' team working on the linguistic aspects of these tribes. Abbi's (2003) work gives a brief grammatical sketch of all the three languages of the Andaman Islands with a vocabulary of basic words. In fact this was the work that gave a fillip to the linguistic study of the languages of these Islands.

The present study was done in almost the same kind of conditions that Manoharan might have found in his visits to Strait Island during 1976-77 and in 1981. Of course, he had better informants at that time and all of the tribes' people could speak the language with comfort as they were just brought from their natural surroundings. We found the same people as our informants this time who were only young men and women at that time. Manoharan (1989) gives a list of the people he worked with during his fieldwork in 1981. Manoharan worked with a population of about 26 (±3) people (1981) out of which 17 are still living as of March 2006. The following table illustrates the age-wise distribution of all the people Manoharan might have worked with and those who are still living.

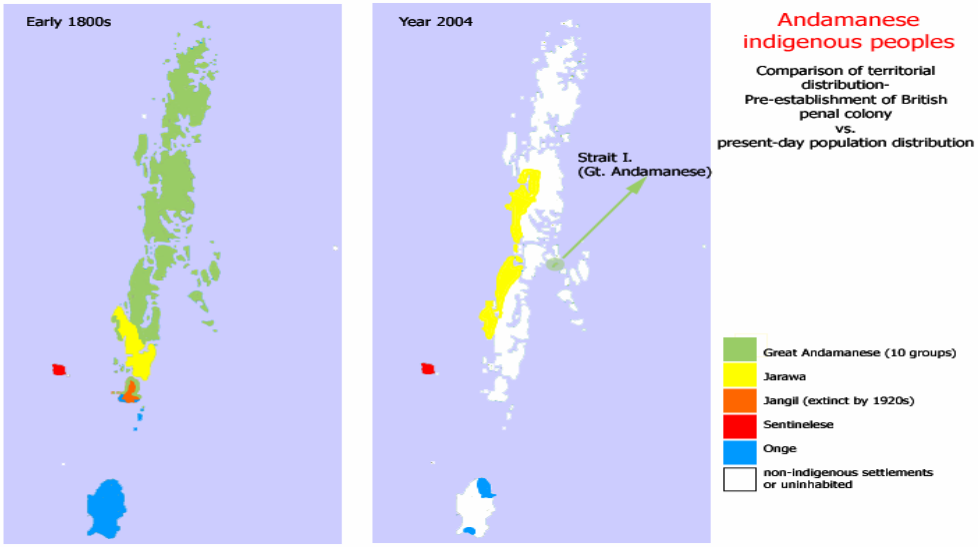
40-60 years	20-40 years	20-5 years	5-0 years
2	6	3	6

Table 2: Age-wise Division of Presently (March, 2006) Living People in 1981

Of these people one can presume that he had talked to at least 2 of the people above 40years i.e. Boa Sr. and Boro. He might have taken data from Lico, Nao Jr., Surmai, Nu, Peje and Look, all of whom were above 20 years at that time.

The Great Andamanese as it is at present represents only the North-Andaman dialect of Jeru with possible mixture of Cari (or Sare as they call it now), Kora (or Khora) and little of Pucikwar (or Pujjukar). It cannot be proved at this time as to how much of mixed elements have emerged in Jeru from other dialects until a comparative study is undertaken which is possible only if one relies on the syntactic data provided by Portman in his *Manual*. About the present situation of the Great Andamanese, Abbi says, after Siegel (1993), "it appears to be close to …a type of Koine" (Abbi, 2006 pp.38). That Cari or Sare bears the most structural and lexical similarities with the present dialect of Jeru is only obvious as both the dialects are from the same geographical region i.e. North Andaman.

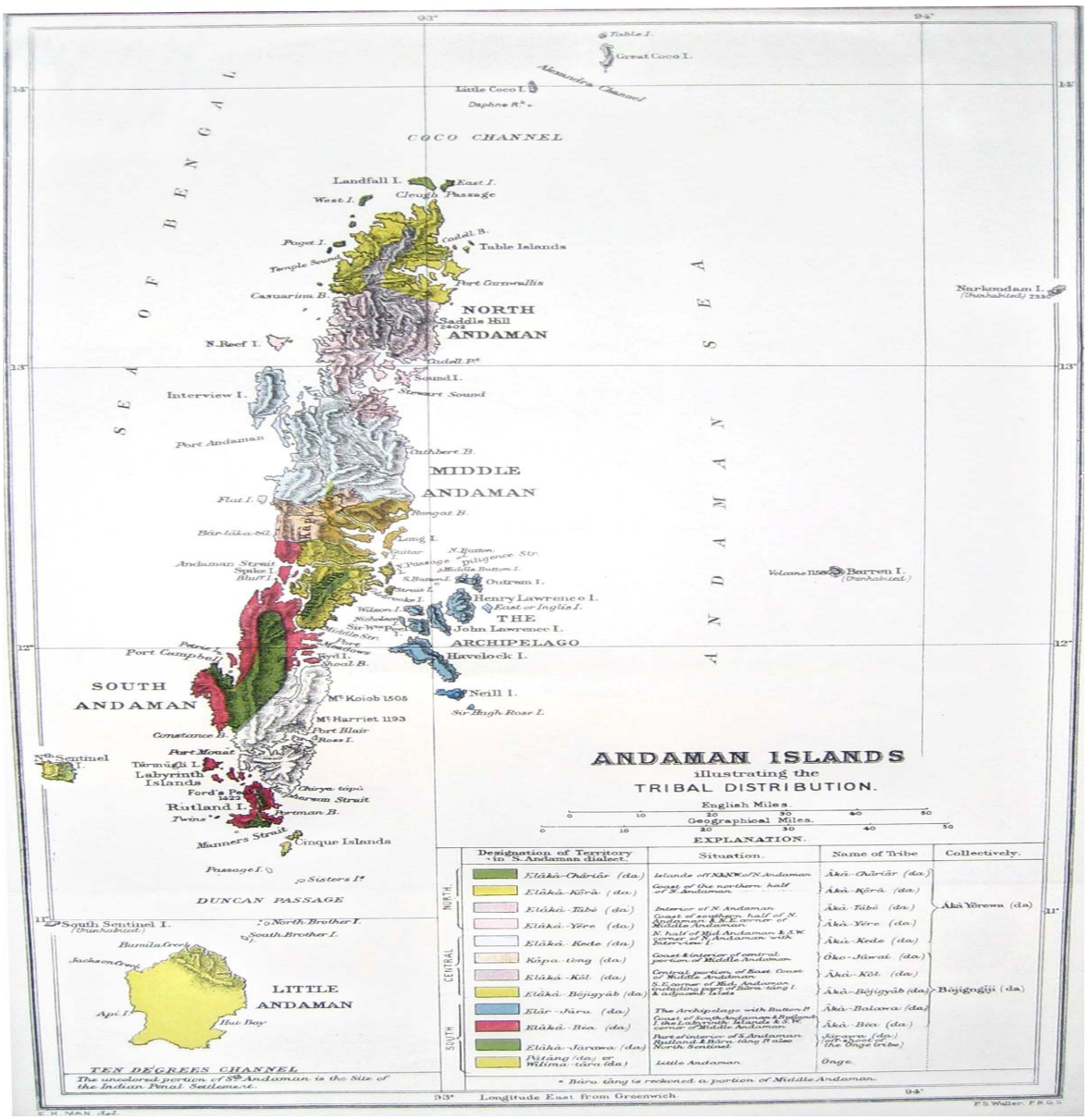
For a quick reference, the following two maps can serve as a guide to the present situation of the language as well as provide a comparison with the situation in the past.



Map1: A Comparative Cartographic Study of the Andamanese Linguistic Spread in a Century

Source: http://commons.wikimedia.org/wiki/Image:Andamanese_comparative_distribution.png

Viewed and downloaded on 17th April, 2006



Ethnolinguistic map of the Precolonial Andaman Islands

From E. H. *Man's Dictionary of the South Andaman Language*, 1923
Source:<http://upload.wikimedia.org/wikipedia/en/f/fd/Andaman_tribal_%26_linguistic_map.jpg> Viewed and downloaded on 17th April, 2006

1.3. Why Computational Framework for GA

The question of why at all working on the computational morphology of Great Andamanese, a language that will have no speakers left say just in ten years is a tricky question to answer. In a situation when more flourishing languages of the world are still to get the helping hand from the technology, working on the NLP of an endangered language may look like a rather too ambitious project. Great Andamanese being an endangered language, it needs to serve an example of how a traditional system which had been alive for thousands of years representing one of the oldest people living in the same conditions, has undergone the threat of a total extinction. This has happened within a century and the speed with which the present situation has taken place is worth noting. If this is taken as an example, one would know the urgent need of documenting and spreading the information about such languages and linguistic situations to the world so that they become aware of the loss of cultural and linguistic wealth and do something to the effect that such knowledge and knowledge systems are kept alive for the posterity. Documenting the language digitally, however, does help achieve this goal in some ways as it serves the purpose of archiving whatever is taken for conservation. But, simply archiving the knowledge in a static way does not give much impetus to the general people to explore or look at it with interest. If there is some dynamic method of archiving which can make exploration easier for the people, this will be of great importance to the people belonging to the community and also to the researchers and, as Himmelmann (2006) suggests, to the discipline of descriptive linguistics, linguistic typology, cognitive anthropology etc. A good reason to the need of documentation and emphasis on the endangered language is given in the books that came in recent years by authors like Crystal (2000), Grenoble and Whaley (1998), Bradley and Bradley (2002).

However a computational framework for a POS cannot be called documentation, or archiving. It can go in one of the two directions – generative or analyzing. The work undertaken here is that of a computational analysis framework for GA verbs – or computational verb morphology in reverse. The good use of this for Great Andamanese could be a quick mechanical interpretation/tagging of the GA verbs, text

understanding, translation from Great Andamanese to other languages etc besides bringing the GA verb data 'in motion' before everybody, everywhere.

The goal of computational linguistics is to "build computational models of natural language for its analysis and generation. First there is technological motivation of building intelligent computer systems such as machine translation systems, natural language interfaces to databases, man-machine interfaces to computers in general, speech understanding systems, text-analysis and understanding systems, computer aided instruction systems, systems that read and understand printed or handwritten text. Second, there is a cognitive and linguistic motivation to gain a better insight into how humans communicate using natural language (NL).

The tools of work in NLP are grammar formalisms, algorithms and data structures, formalism for representing world knowledge, reasoning mechanisms, etc. Many of these have been taken from and inherit results from computer science, artificial intelligence, linguistics, logic and philosophy".[2]

The program being undertaken here is that of pertaining to an area of NLP. The dynamicity of the program is of greater value and importance. The program being produced through this endeavor is limited to a morphological analyzer only. The morphological analyzer presented here can serve as a basis of further analysis of the data. The whole of the text can be analyzed if some new modules are added to the present program. Besides serving as an online text analyzer for Great Andamanese, the web thus produced will give a greater visibility to the language and the community.

[2] Bharati, A. et. al. 1999.

1.3.1 Computational Morphology and POS tagging of Indian languages

1.3.1.1 Areas of R & D under Computational Linguistics

Computational linguistics started first as an MT project, however, it soon reached new areas where it contributed a lot, both to the area of linguistics and computer science. The major areas of research and development in computational linguistics can be summarized as below.[3]

- **Corpus Building**
- **Machine Translation and Machine aided Translation**
- **Speech Processing**
- **Localization**
- **Human-Machine Interface Systems**
- **Tools**
- **E-Contents**

Corpus building is one of the basic needs of language processing. A good corpus, properly tagged and annotated, serves the basis of a good and efficient linguistic tool. Corpus data are the raw fuel for many applications and the test bed on which NLP applications are evaluated. The field of corpus building has become so vast that a whole disciplinary term of corpus linguistics[4] has emerged and at several places this being taken branch of academic endeavour complete in it. Several corpora have been built over the years in many languages for example the British National Corpus and the Bank of English, the European Union funded multilingual English-French-Spanish corpus, the CLAWS English text tagger etc. In India, the same effort has been on for Hindi and other languages at several places for example, IIIT Hyderabad, IIT Kanpur and Delhi, MGIHU Wardha etc.

---

[3] A good summary of language technology in India has been given in Jha's article in *Currrent Trends in Indian Languages Technology* published in *CSI Communications*. This article has served as guide for this section and for many other resources given throughout this work.

[4] (*See* Botley & McEnery, eds. 2000, Mitkov, T. 2003, Sinclair, 1991 etc.)

The history of **machine translation** is traced from the pioneers and early systems of the 1950s and 1960s, the impact of the ALPAC report in the mid-1960s, the revival in the 1970s, the appearance of commercial and operational systems in the 1980s, research during the 1980s, new developments in research in the 1990s, and the growing use of systems in the past decade. However, reaching a hundred percent correct output of a machine translated text is still to be achieved and research and development in this area is still on to bring about the best that a machine can deliver. There are several MT systems already on sale by private companies that one can find on the web, each claiming better results. In India, the initiative towards work on MT was led by IIT Kanpur. Its AnglaBharati and AnuBharati projects are ambitious efforts towards developing an MT system not just for Hindi and English but for also other regional languages.[5] Similar kinds of efforts are also on at IIIT Hyderabad (the Shakti MTS), C-DAC Mumbai (the MaTra MT system), C-DAC Bangalore (the Mantra project) and at IBM IRL.[6]

Speech processing is the study of speech signals and the processing methods of these signals. The signals are usually processed in a digital representation whereby speech processing can be seen as the intersection of digital signal processing and natural language processing. There are several sub-categories in this area of COLING where research and development is on. This area seen many successes for example speech recognition, speaker-recognition, speech-synthesis and so on. Besides these, of particular importance is the text to speech (TTS) and speech to text (STT) systems that have come in several languages.[7] Efforts in this area in India are made mainly at TDIL, C-DAC centres, Media Lab Asia centres and at Tata Institute of Fundamental Research (TIFR).

---

[5] For more on these systems and an online version of the translation system, visit: http://www.cse.iitk.ac.in

[6] All these institutions and agencies give details of these projects and the R&D work on their respective websites.

[7] See Jurafsky et al., 2000.

Localizing the software applications and platforms are the needs of the present day world so that the great digital divide that has created due to linguistic barriers could be bridged. The linguistic diversity is considered as an asset for the world and there has been a danger on it as English and its resources are gaining greater grounds in several areas and thereby limiting the use of the other tongues for some specific domains. Several governments have understood the need of it and a remedy is eagerly sought through localization of software applications and platforms. Many public and private endeavours in India are being put forward in the area of localization for example Indix, a localization of graphical user interface of Linux operating system being developed at NCST through TDIL program of Ministry of Information Technology and Telecommunications, Govt. of India.

Tools are other areas of computational linguistics where research and development give short term goals that serve major needs. Some examples of tools developed and under development process for Indian languages are the ITERM- an Indian script terminal for Unix X windows, word processors, Anusaarka- a language accessor among Indian languages to Hindi, Sanskrit Authoring Systems, Devanagari search engines for Unicode etc.

Linguistic e-contents are another widely used area of computational linguistics. There are several such contents already available on the web that can be availed freely. The bilingual electronic dictionaries are available in many languages. Word nets, corpora, speech databases, web based education systems are some examples of linguistic e-contents.

1.3.1.2 History and Background of Computational Morphology

Computational morphology is a branch of computational linguistics. Computational linguistics itself is a relatively newer branch of intellectual endeavour that saw its emergence in the ALPAC report in the USA where it was proposed that the ‘machine

translation be abandoned as a short-term engineering goal in favour of more fundamental scientific research in language and language processing' (Kay, 2003). Computational linguistics is an interdisciplinary field concerned with the processing of language by computers. Today, the area of natural language processing has emerged as a vibrant field where work is on in developing computational tools in many languages of the world. Already the NLP tools for major languages are of great help to many people and it also helps new people to the language community. Although, languages with lesser resources are either struggling to find a place in the computational architecture on the whole or are facing the threat of getting eliminated as its own speakers are shifting loyalty to the community that have a better resourced linguistic system, the language this dissertation is based on being an extreme case of this kind.

Computational morphology is nothing but implementing what linguistics finds in the morphology of a language. The most basic task in the area of computational morphology is to "take a string of characters or phonemes as input and deliver an analysis as output" (Trost, 2003).

The morphology of a natural language is complex enough to have several rules to constrain a set of lexicon that have a potential to give an analysis computationally. In processing of a natural language, the lexicon plays an important role followed by the morpho-phonological rules that govern the lexicon. The lexicons may be of different kinds depending on the demand of the language structure.

1.3.1.3 Current Status of Computational Morphology

A canonical form of lexicon is called a ***lemma*** lexicon. It takes usually the base form "as the representative for all the different forms of a paradigm."(ibid) The *lemma* lexicon may be suppleted with other lexicons comprising of affixes and the relevant morpho-phonological and morpho-syntactic information about how they are concatenated with the *lemma*s. There are possibilities that there will be forms that will

not conform to any of the morpho-phonemic and morph-syntactic rules. To account for them, one needs to have some exception handling mechanism as well.

The mechanisms used currently for the computational morphology are divided in two groups- **Finite State** and **Alternative methods**. The finite state mechanism has the potential to account for most of the morphological phenomena. However, it is not possible to describe the non-concatenative phenomena (like vowel harmony, root and template morphology, reduplication etc.) with only a finite state automaton. The **two-level morphology**, deriving its ideas from generative phonology (Chomsky and Halle, 1968) thus takes care of morph phonology. The alternative methods "include so-called paradigmatic morphology described in Calder (1989) and the DATR system" (Evans and Gazdar 1996). Common to both is the idea to introduce some default mechanism which makes it possible to define a hierarchically structured lexicon where general information is stored at a very high level. Lower in the hierarchy this information can be overwritten. Both systems seem to be more concerned with morphosyntax than with morphophonology. It is an open question if these approaches could somehow be combined with two-level rules. For more on the computational morphology one could see Sproat (1992), Black et al. (1992) and Spencer and Zwicky (1998).

The data used in processing of a natural language needs to be structured properly to suit the computational model required for the language. As human languages have a variety of linguistic structures through which they express themselves, to capture those structures computationally a suitable framework is developed. What follows is brief introduction to the currently applied techniques in storing data in a structured manner.

Text segmenting is the first thing that needs to be done before any processing of the text takes place. In this part generally the linguistic units like words, phrases, punctuation, numbers, alphanumeric etc. are put separately so that when the input is given, it automatically recognizes these segments and prop them for further analysis.

The next step is that of **tagging** or **tokenization**. All the parts of speech are usually given a tag that contains information on the meaning, function and use of the element. The POS taggers of this kind are used for several purposes such as serving as a pre-processor to many NLP automata. Large tagged text corpora (e.g. British National Corpus; Bank of English Corpus) are used as data for linguistic studies. Information Technology applications, e.g. text indexing and retrieval can benefit from POS information for example nouns and adjectives are better candidates for good index terms than adverbs, verbs or pronouns.

Speech processing can benefit from it e.g. the pronoun *that* in English is pronounced differently from the conjunction *that*.

1.3.1.4 POS Tagging for Indian Languages

Much work in the area of POS tagging in India has been carried out and is still on at several places and in several languages.[8] Of particular mention are the works carried out at Centre for Development in Advanced Computing (CDAC), IIIT Hyderabad and IIT Kanpur and JNU New Delhi. CDAC and IIT Kanpur have been very instrumental in developing many tools for Hindi in particular. The machine translation (MT) project is running at these centres. Besides the works at the institutes and government funded agencies, several private units are also putting their effort in the area of NLP. Here it would be worthwhile to list some of the major players in this field who are the implementing agencies of TDIL programs.

1. IITs at Kanpur, Delhi, Mumbai, Chennai, Kharagpur, Guwahati, and Roorkee
2. IIIT, Hyderabad
3. National Centre for Software Technology (NCST)
4. Centre for Development in Advance Computing (C-DAC)
5. Indian Statistical Institutes

[8] The journal on language technology in India *Vishwabharat*, published by Ministry of Information Technology, Govt. of India, keeps an eye on all the new research and developments in this area and reports them.

6. Utkal University, Bhubaneshwar
7. Thapar Institute of Engineering and Technology
8. University of Hyderabad
9. Indian Institute of Science, (IISc) Bangalore
10. Central Electronics Engineering Research Institute (CEERI), Pilani
11. Tata Institute of Fundamental Research, Mumbai
12. Anna University, Chennai
13. Vanashtali Vidyapeeth, Banasthali, and
14. Jawaharlal Nehru University, New Delhi

Besides these, there are several private enterprises underway in this direction, the majors among them being IBM, Microsoft India, Tata Consultancy Services, HP, HCL, Webdunia etc.

POS taggers are also being developed at small centres and institutes for regional languages by individual and team efforts. For example, an analyzer being developed for Manipuri and a tagged text corpus developed from using the web as source of data in Bengali at Jadavpur University, Kolkata. A similar kind of effort has been made for Assamese and Manipuri at IIT, Guwahati. A morphological analyzer is also developed for Kashmiri at IIIT, Hyderabad. The main thrust in software applications has been Sanskrit as it is believed that Sanskrit is the base of all the major Indian languages and once it is done, it will automatically lead to the solutions of other languages. Sanskrit informatics has been going on at C-DAC, Jawaharlal Nehru University, Tirupathi Sanskrit Vidyapeeth, Vanasthali Vidyapeeth, Academy of Sanskrit Research, Melcote, Mysore and other places. Sanskrit informatics has been a pet subject to many of the institution and organizations. Jha and Mishra (2005) have done a comparative study of the work in Sanskrit morphology and have proposed an alternative model that claims to work better than others.

1.3.2 Computational Framework

The computational framework for the task brought about here is a very simple one except for the fact that it applies some rules with a linear paradigm for the verb morphology. The grammar is formal and context-free, based on the researcher's typological studies. It does not claim for a hundred percent correct results because all the levels of grammatical adequacy have not been implemented.

1.3.2.1 What is Computational Framework?

The term "computational framework" as used in the title of the dissertation was chosen on account of an uncertainty about the results that the researcher might have found after the fieldwork. However, that framework is now clear as given below and in the fourth chapter. A computational framework may stand for, to begin with, a single corpus containing some words of all the linguistic information to serve as template for the whole of a grammar of a language or just a part of it. Thus, a framework may consider only phonology of a language or it may take morphology, morphophonemics, syntax, stylistics, semantics, discourse information and so on also in consideration, depending on how robust and how efficient the system is required to be.

1.3.2.1.1 Understanding the Task

For any task, it is necessary to first understand it. It is easy sometimes to classify language into grammatical levels of phonology, morphology, syntax etc. This classification is more useful for academic endeavours than for practical applications and implementations. To implement a computational rule, one needs to understand it in the perspective of its position in its entirety. For example, a phonological rule may influence other levels of grammar such as morphology, morphophonemics etc. so much so that it may be difficult to precisely demarcate the division lines between two grammatical levels. Understanding a task implies to understand its position vis-à-vis the whole structure of the language under consideration.

The next step after understanding the task is to ponder over its **viability of getting handled by machines**. Proper understanding of the task may give several perspectives over its solution. It is here that one has to think of a way how this task is to be handled by machines. Obviously, machines understand the language of valid logic and not all kinds of logic can be applied here. One has to think in binary terms and get to the root with the initial threads of binary approach, step by step.

When it is understood that a task is feasible to be handled by machines, one has to look for ways to implement it. The process of reaching a solution may take time and it may involve first an initial **algorithm** or **pseudo code**. Writing an algorithm for a specific problem gives a fair idea of how the task is to be implemented. Algorithms give a formal view of the solution before it has been implemented and tested. After **formalizing** the solution one has look for ways of how the solution is brought about i.e. the **delivery and display methods**, decisions on what **tools and techniques** to use and so on. For example an application may be displayed in several ways such as on command-console or it may be a graphical, user friendly interface. There may be several tools and techniques available; one has to take decisions on what is best suited for the specific purpose of the developer.

When all of this has been done, finally one comes to the **implementation** of the task. The implementation process is the most crucial as all the planning done before comes to use here. After the implementation of each of the modules, one needs to check it at all levels. **Testing** the program is another important step towards finalizing the endeavour. Every entity in the world has a life cycle; programmes are no exceptions to this. Every program has to be updated after a period of time and the updating or **upgrading process** should be regular, depending on the newer requirements.

1.3.2.2 Brief Description of Problem and the Solution

The problem undertaken here is that of belonging to the NLP area. We are dealing in between some input and output. In the midway there has to be the whole process of

the analyzing, segmenting, formalizing, formatting and so on. The problem is that of imparting the understanding of a natural language to machine through available means. The goal is to identify the verb phrases in a linguistic text given as input and analyze it by segmenting the phrase into smaller linguistic units and giving a proper tag for each of them that carries their linguistic content in the phrase. In the case of Great Andamanese, it involves implementing the verb paradigm that contains how prefixes and suffixes are conjugated to a verb root.

Although, the structure of the paradigm is linear, it has some rules that require the affixes to be in particular order. There are cases of optionality, obligatoriness, constraints and choice of meaning depending on the typological context of the occurring element. The following block diagram illustrates the solution given for the problem. A detailed study of this can be found in the fourth chapter.

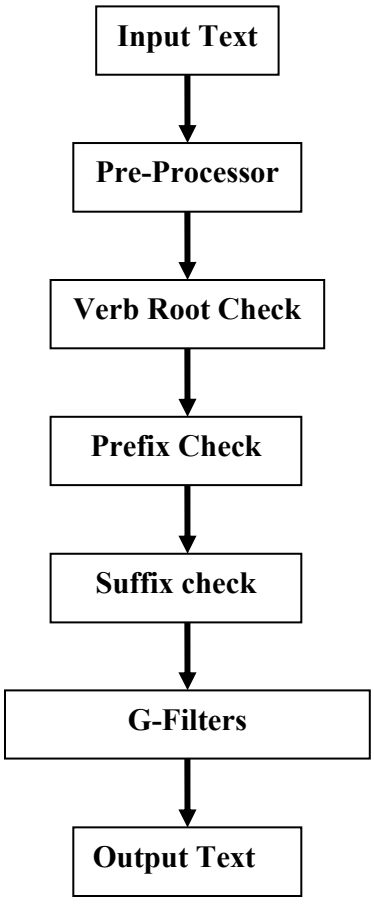


Figure 2: A Block Diagram for the Structure of Great Andamanese Verb Analyzer

1.3.2.3 A Brief Introduction to the Program/Technology Developed

The name of the technology developed has been give as Great Andamanese Verb Analyzer or GAVA for short. It analyzes the verb phrase from a given text in Great Andamanese. The nature of the technology developed is that of a knowledge system which can also function as an aid to machine translation and other NLP applications. In other words, it can be at the levels of a product, a technology and as a sub-system. The program is an online Java/JSP application running on Tomcat-Apache webserver. It has a JSP front end and a text files as back-end. It can be hosted on any platform and can be viewed as HTML. The input and output is in IPA UTF-8 (refer to the Appendix I).

This program has been developed as a web application using Java/JSP on Apache Tomcat platform. As the language does not have any script, IPA characters provided by Lucida Sans Unicode font have been used. Any Unicode font supporting IPA can be used to view and input the texts. However the technology is fully portable from one system to another as only text files for database has been used and the tools used are platform independent.

The program is highly scalable and can be easily extended to work as text processor and tagger for Great Andamanese or languages with structural similarities. Most of the documentation for the program part has been completed. Testing of the software is currently undergoing and we are getting more than 90% results. An appendix (IV) to this effect has also been attached showing the result of the most frequent verb phrases in the basic sentence list.

1.4 Methodology

The dissertation is primarily based on the first-hand data that I collected for Vanishing Voices of the Great Andamanese (VOGA), a SOAS project of Endangered Language Documentation Program on Great Andamanese. The data were collected during October, 2005 and March, 2006. The extensive fieldwork concentrating

mainly on the linguistic aspects of the Great Andamanese community was carried out in Port Blair where some of the members of the community reside for their livelihood and other necessities and in Strait Island. In addition, secondary sources in the form of written documents were also used to elicit relevant data.

1.4.1 The Informants and the Field Situation

A word must be said about what it was like working with an endangered community as it also takes on how a linguist performs in the field. Working with an endangered community was quite a revelation. The social norms that one might want to see in the community that had been isolated for thousands of years had almost taken a vault face.

In a community comprising of about 50 souls, only six or so could claim to speak their native tongue. It was difficult to get a native speaker for a nice sitting. Not only this, even among these last speakers, the level of proficiency in their own tongue has deteriorated so much that none of them could say that they are at their best in speaking their own language. All of these terminal speakers complained that they were forgetting their mother tongue for want of use of the language in their daily life as all of them had shifted to Andamani Hindi.

In this kind of a situation, if one sat for an hour, the output would be lower than what one would expect from a native speaker normally. There were several times when the informant got stuck with a word or was not sure about a construction. It was fortunate that I had an access to the earlier works carried out on this language after their resettlement.

Before going to the fieldwork, I had already worked on the secondary sources available in the language. I had developed a digital database on this language that was based on Manoharan's (1989) work and the works carried out by Abbi and her team in 2001. Besides these, I had also had the chance to listen to the tracks recorded in the same project by an earlier team of the project VOGA that stayed there for a month.

Thus, I was already a bit tuned up to the language and the community before I started for my rather long stay in the field.

Our main problem in the field was getting the permission to meet our informants. As the Great Andamanese have been declared a Primitive Tribe by the Govt. of India, the outsiders are not allowed to associate with them. It is only with the permission from the Administration that one can meet with them either in Port Blair where a guest-house named Adi Basera has been built to lodge the primitive tribes or in Strait Island, 63 nautical miles away from the city of Port Blair. The area where these people reside has been declared as the Tribal Reserve. So the first thing that we had to do was to get permission. And this took our two precious months during which several communications underwent up from the Ministries down to the Offices. Much of our energy was spent moving around the administrative premises in Port Blair which always turned us down. By the time we finally got the permission in December, the tribal community had already known about us and our interests. There were some who had already started living independently out of the Tribal Reserve area of Adi Basera and had taken houses on rent or were living in the Government quarters along with other employees in the township of Port Blair. So when we first went to Strait Island, we had these people to give us company to their belittled land.

As I said earlier, extracting linguistic data was the most difficult thing that one could do with these people. They did all the other things with much enthusiasm except for sharing something in their own language with us. It seemed that they wanted to take more than to give, realizing that they had nothing to give to the world that it will receive with respect.

So the task uphill was to tune them to go down memory lane and think with a sense of pride so that they be in a good mood to share something with us. This was a part of forming an ideal condition to extract linguistic data. If there were interruptions, the informant would give a wrong sentence which I gradually started inferring and at that time it needed either to be cross-checked or move forward to keep the informant's

composure intact. This way I had to check the same sentence twice or thrice or even several times until I obtained the right answer.

And this was not an easy task. Given that all of our informants were prone to alcoholism, getting them to sit in a good mood was always a difficult task. A constant demand from all the members of the community was for alcohol that could topple them up. Whenever, they got a chance, they would go for it. So much so that even for the interview sessions, they would demand the same which unfortunately we were not in a condition to provide as it would mean a violation of our work ethics. I myself being a teetotaller, most of them always found me a boring character. Fortunately, I was still liked by some, including my main informant, Nao Jr., who, although was prone to drink, was docile enough to keep himself under control and when needed also abstained himself from drinks. Otherwise, I was not a good character in their eyes as I did not approve of their drinking habit for which they thought I was saving the money that was supposed to go to them. But in the end, despite my brief fracas and shows of temper, I emerged a good man in their eyes and was loved by all.

1.4.2 The Bilingual Elicitation Method

Our method of eliciting the data was bilingual as all our informants spoke what is termed as Andamani Hindi. Andamani Hindi has emerged with the emergence of the settlement in the group of Islands. Andamani Hindi is spoken by the settlers of different linguistic origins such as Bengali, Tamil, Malayalam, Telugu and other communities from Burma and Jharkhand, besides small communities from other parts of India. As not many people from the main Hindi Heartland of India is settled there, the lingua franca that emerged there has got a shape of its own to which the contribution came from all the speech communities settled there. Thus the Hindi spoken by the Great Andamanese people serves only an indicating help and not the actual translation. We had to devise our own method to get at the actual construction. So what I did was to start speaking in Great Andamanese based on the assumptions formed on the basis of the earlier data I had an access to. And this method worked as I found myself groping inside the structures of the language with much improvement

coming day by day and later so much so that I also got comments to form sentences of complex kinds on my own as they could not make it or were not sure about this.

1.4.3 The Data

All my data procured here are well annotated and recorded. I have given examples of data after having attested it and having recorded it. All the evidences proving my findings and analysis are based on the data that have been recorded and rechecked with the native speakers and I have not given any sentences of my own construction to validate my analysis. Writing this is necessary as, the language being endangered; it is easy to influence the speaker to give their nod for a particular kind of sentence that may not be correct (which was a case that we found when we had to face an enthusiast researcher working with the same community). Also, all the data taken has been recorded with all the paralinguistic information available in the tracks. Although, the interview tracks grew much due to this method, I believe it has contributed to the quality of the interview period.

There were several gadgets and equipment used to ensure the quality of data. The following is a list of the equipments we used to elicit audio data:

1. Sony Mini Disc Walkman cum Recorder set with headphones
2. Sony Mini Discs as the medium of temporary storage
3. Sony hi-quality microphones

Besides these, some times I also recorded data in the Sony Memory Voice Recorder that uses a memory chip to record data in the MSV format. It is notable that the Sony Mini Disc walkmans the devices that store data in the WAVE format hence, the quality of the audio generated is the best. However, one can change the format of these audio files into other formats as well.

The Leipzig glossing rule has been for the annotation of the transcribed data. But new terms have also been used wherever necessary such as CLSM etc.

The transcription for each track recorded was done by me both in the field itself and afterwards for cross checking. The five months time spent with the community has also added to the precision of the transcription. Other necessary tools like the Praat analysis of the sound segments are yet to be done.

1.4.4 The Analysis

The analysis of the data is done thoroughly with the help of the native speakers' intuition and crosschecking method. Abbi's *Manual* (2001) has been of quite a help in this regard. Wherever necessary, I have taken resort to statistical method, making lists of the occurrences and checking their frequencies to come at general conclusions.

1.4.5 Computing the Verb Morphology

After the verb paradigm has been formed in the third chapter, the main purpose for the computational framework is to implement it i.e. to make the machine learn it and analyze the verb phrase and tag the functional parts in it. The verb paradigm and the associated rules have given the raw algorithm for the computational framework suggested here. On the whole, the framework is based on this paradigm and at points also falls short of the goal as described in the fourth chapter. The paradigm being mostly structural in approach, no morphophonemic rules have been applied and the system relies on the lexicon and the reordering cum positional rules inherent in the paradigm.

For an access to a wider mass of people, the program has been web enabled and uses Java/JSP technology web served by the Apache-Tomcat technology. JSP is its front end and the data files in text format are its back end. The interface is thus user friendly and easy to use.

1.4.6 The Chapterization

The dissertation is divided in five chapters. The first chapter i.e. the ongoing one, deals with the introduction to the problem ahead and gives a fair deal of information about the land, the people, the language and the problem of making a computational

framework for the morphology, besides giving an account of the use of the language technology in the field of linguistics in general and morphology in particular.

The second chapter is focused on giving a brief typological sketch of the language. It also discusses in detail about the linguistic work done earlier in this language.

The third chapter is devoted to verb morphology and here I am giving a detailed analysis of the verb, its inflections and the verb phrase. By default, I also had had to discuss other areas such as the sentence structure, the pronominal prefixes which act as subjects in the small clauses and the object clitics that form a part of the verb phrase. On the whole, this chapter constitutes the main topic of the research and based on the analysis of the data collected by this researcher, posits a solution for accounting the verb morphology in the language.

The fourth chapter of the dissertation gives the details of the computational framework suggested for analyzing the verb phrase of Great Andamanese. In this chapter, the structure of the Great Andamanese Verb Analyzer has been given in detail. It discusses the framework proposed and implemented as well as the performance of the program.

The fifth chapter is the conclusion part which takes on all the work done through the dissertation and gives conclusive remarks on them.

Chapter 2

A Brief Typological Sketch of Great Andamanese

The earliest record of any work on grammar is that of Portman's (1887), where he gives an 'Introduction to Grammar' that is for all the four languages of the *Manual.* Though this is only a brief, generalized sketch of grammar spread in 8 small pages, often not on the standards of modern discipline of linguistics, the content in these 8 pages are of great importance as it gives some generalized statements on the four languages of Andamanese group. In these four languages, Jeru is not under study. However one can find several similarities between Jeru and the two North Andaman languages of that time i.e. Aka-Chariar or Sare and Aka-Kede or Kede, both of which are extinct now.

The eight page introductory grammar of Portman is divided in four sections dealing briefly with the following topics:

1. Alphabet,
2. Word Formation,
3. Parts of Speech and
4. Syntax.

Manoharan's1989 book gives a fair deal of Andamanese structure as it was after their resettlement in Strait Island. He has a fair deal of work on phonology, noun morphology, and verb morphology of the language. The last work on the language and linguistics of Great Andamanese was done by Abbi in 2001 when a pilot survey was carried out on all the languages of the Andaman Islands except for the Sentinelese. What follows is a review of all the works carried so far.

2.1 Phonology

The alphabet section in Portman (ibid) deals with the representation of sounds in Roman script and its pronunciation. Nothing of the phonetic inventory is given as is wont in modern linguistics. Manoharan devotes one whole chapter on phonology of Great Andamanese. Manoharan gives a list of 31 phonemes out of which he says 3 are suprasegmental phonemes, 28 segmental phonemes. According to Manoharan, there are 21 consonants and 7 vowels in this language. The suprasegmental phonemes in Great Andamanese according to Manoharan (ibid pp: 29) are length, accent and juncture. Intonation in Great Andamanese is phonetic in nature. Compared to that of Abbi's (2006) observations, we get a list of 26 consonants and 7 vowels sound segments. But Abbi's findings are based on a list of phonotactics and not on minimal pairs. Let's discuss the phonemic inventory of this language in some detail.

2.1.1 Consonants

Manoharan (ibid pp: 11) gives the following chart of phonemic inventory in Great Andamanese:-

Place of Articulation→ / Stricture↓		Bilabial		Alveo-Dental		Alveolar		Retroflex		Palatal		Velar		Post-Velar	
		Vl	Vd	Vl	Vd	Vl	Vd	Vl	Vd	Vl	Vd	Vl	Vd	Vl	Vd
Stop	Vl. Vd.	p	b	t	d			ʈ	ɖ	c	ɟ			k	
	Asp.			t^h											
	Slit	ɸ										x			
Fricative	Groove					s									
Nasal			m				n				ɲ		ŋ		
Lateral							l								
Trill							r								
Semi-Vowel	W									y					

Table 3: Phonemic consonants in Great Andamanese (adapted from Manoharan, 1989)

Compared to the above table the following table (adapted from Abbi, 2006 pp: 2006) illustrates the consonantal sound inventory of Great Andamanese.

	Bilabial	Labio-Dental	Dental	Alveolar	Retroflex	Palatal	Velar
Plosive	p b p^h		t d t^h		ʈ ɖ $ʈ^h$	c ɟ	k k^h
Nasal	m		n			ɲ	ŋ
Trill				r	ɽ		
Fricative	ɸ β	(f)		s		ʃ	(x)
Lateral	l^w			l			
Approximants	w					y	

Table 4: The Consonant Sounds of Great Andamanese (adapted from Abbi, ibid)

Out of these sounds, there are several which come only in particular cases and may not be phonemically distinct. The sounds that include in this list are /ɽ, β, ɸ, f, x, l^w, y, and w/. Special mention must be made of one of the informants named Peje (58yrs circa 2006) whose phonemic inventory has some 'aberrations' as he frequently uses fricatives where other speakers use plosives. This aberration was noted in Abbi's fieldwork during 2001 and also given as peculiar sounds in Abbi (2006) where it is clearly mentioned that the following sounds are in 'free variation at the intra-community level i.e. within the same clan'-

[ɸ ~ p^h ~ f]
[β ~ l ~ w ~ l^w]
[k^h ~ x]
[s ~ ʃ]

Even the retroflex trill [ɽ] is in free variation in the same way with the alveolar trill [r]. The phonemic status of the approximants again is not clear. However, Manoharan (ibid pp: 21) finds /w/ occurring medially and finally and /y/ occurring in all the three positions, Abbi (ibid: 43) finds these two sounds occurring in all the three positions. Thus, if we take into account all of this there will be a total of 22 phonemes, including the approximants. This is just one more than what Manoharan gives in the chart above. However the difference of sound segments thus comes as follows.

First of all, the /k/ sound to Manoharan is post-velar and not velar; however its fricative counterpart he says is velar, which he says is a groove sound. Another peculiar phoneme he reports of is the bilabial slit /ɸ/. We could not validate this sound, rather what was found to be phonemic was its variant aspirated. In fact, as no phonemic analysis of the language is done in this dissertation, it is bound to build on what the others have said. What was found in the data elicited was, like most of the Aryan languages of India, continuum of the aspiration feature in the voiceless plosives at the junctures of the bilabial, dental, retroflex and velar with just one exception at the velar position. One does not know whether this is a recent influence of Hindi used by the speakers there or was there from the earliest.

2.1.2 Vowels

There is a consensus on the number of vowels used in the language and these are the following:

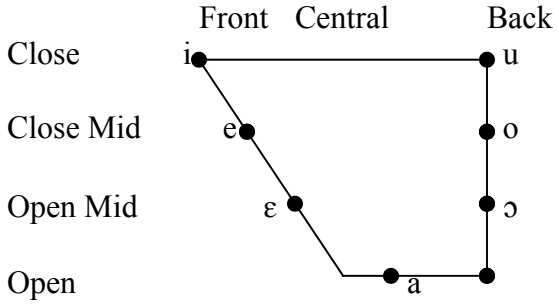


Figure 3: Distribution of Vowels in Great Andamanese

The vowel graph shown above is corroborated by both Abbi and Manoharan. However, both of them do not use the terminology used up there i.e. they use the terms high and low instead of close and open to describe the tongue position during vowel utterance.

2.2 Morphology

In the word formation section, of particular importance is Portman's remarks on verbal prefixes. According to Portman, the verbal prefixes like /-ot/ and /-on/ merely

serve the purpose of intensifying the verbs and ‘care should be taken not to confound them with the pronouns used as prefixes (Portman, 1887 pp.:2)’. This is a statement that is of direct relevance to the findings and analysis of both Manoharan (1989) and Abbi (2003, 2006). The verbal prefixes are posited as ‘clitics’, which is also validated by my data and go against Portman’s statement that they are merely intensifiers. Though, no comparison can be done between the two works as the former lacks the validation by analysis and proper evidences.

While the compound word formation examples given by Portman are fine, his statement again about the ‘particles’ like ‘b’, ‘t’ and ‘l’ as serving just the cause of euphony will not be acceptable to modern linguistics.

2.2.1 Nominal Inflection

Regarding structure of nominal element, Manoharan gives examples of nouns from mono-morphemic to tri-morphemic to ploy-morphemic (1989 pp.:59). As I have not yet undertaken any analysis at the syllable structure level, I cannot say anything on it. However, it is clear that nouns are not marked for number and gender. (Portman, 1887. pp.:4) Nouns are case marked with suffixes. Portman (ibid) says that case is also not marked, which is not correct. See the following examples:

1	GA	ram konabi ɟiyom				
	MB	ram	kona-bi	iɟi-om		
	Gl	Ram	tendu-ACC	eat-NPST		
	Egl	Ram is eating a tendu.				
	Src	Nao Jr/ 15.11.05/ MD1 Gr II T3/ Narayan C-1				
	Rem	**SIMPLE SENTENCE**				
4	G.A.	aʃyame bəs kuttaral konabit beliŋo				
	MB	a-ʃyam-e	bəs	kotr-al	kona-bit	beliŋ-o
	Gl	SPEC-Shyam-	bus	inside-LOC	tendu-ACC	cut-PST
	Egl	Shyam cut the tendu fruit in the bus.				
	Src	Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-1				
	Rem	**SIMPLE SENTENCE/ACCUSATIVE**				

With these two sentences it should be clear that the cases are marked on the noun as postpositions. The case markings on nouns or noun phrases are as follows. However

this is not an exhaustive list of the postpositional suffixes as will be clear later in the dissertation, this gives a fair picture of how the cases should work in the language

CASE	Andamanese Form	Source/Validation
Accusative	bi ~ -i ~-e ~ɸ	All
Instrumental	ta	All
Ablative	ta	All
Possessive	-iʃo ~ ico ~ -ut ~er	All
Sociative	cul	All
Purposive	-ɔkka	After Manoharan
Benefactive	kocara	All
Locative/Directional	-ak ~ -a	All
Locaive/Stative	-al ~ -il ~ -el ~l	All

Table 5: Case Markings on Nouns

Besides these suffixal case markings, there are also postpositions that occur in isolation. However they are all of locative/demonstrative in nature and are constructed through rules as they are all suffixed with the stative/locative suffix /-al/ i.e. they are bi-morphemic constructions which necessarily take –vl [vowel+lateral] representing 'state'. Thus, while the bare morpheme /kutra/ means 'inside', the locative suffix /-al/ gives the structured meaning to it so that it comes in isolation.

Case/English Gloss	Andamanese Form	Source*
on	tutara:l	A-61
inside	kutara:l	A-61
under	tumbol	A-62
near	cul	A-61
away	tacak^{h}ul	A-62
in front of	tɛrcok^{h}el	A-62
behind	tutbɔl	A-63

Table 6: Postpositions Occurring in Isolation
* For source see Appendix II

2.2.2 Adjectives

Portman (ibid) says that there are no degree of comparisons and the 'meaning of a word is intensified by a prolonged enunciation of one of its syllables', which is true. In fact, the language is also very poor in using adjectives which validates a statement made by Portman in his *History of Relations* that these people lack interest in the

beauty spread around in the islands. The degree of comparisons is also not present as they give just the clausal values for the comparative and superlative adjectives. See for example sentences in the BSL for comparative contrastive.

2.2.3 Numerals

Portman states that Andamanese cannot 'count more than two. All numbers higher than two are only vaguely represented' (ibid). While Manoharan reports numerals up to three as follows:

ondoplɔ = one

oŋjikɔ = two

onda:ɸol = three

That of CIIL's primer (1994) gives numbers only up to two as follows:

ontoplɔ = one

taterp^{h}uyi = two

On my asking, the subjects could give numbers only up to two which confirmed what was given in the CIIL's primer above.

The ordinal numbers are also up to only three (Manoharan, '89 pp.:81) out of which I could validate only up to one that of the first as follows:

64	G.A.	unne tujulu ukut^{h}u		
	MB	unne	tujulu	ukut^{h}u
	Gl	3SG-PL.OBJ.CL	early	3SG.SUBJ.CL-reach
	EGl	He reached before them.		
	Src	Nao Jr./15-11-2005/MD-1, G-12 T-3 CIIL Q-4/NarayanA-58		
	Rem	**ORDINAL/ADJECTIVE.**		

Even here, the supposedly 'ordinal number' is in fact temporal verbal in nature, suggesting non-existence of the ordinal numbers.

2.2.4 The Importance of Pronominal Prefixes

Andamanese has a very complex and enriched system of using the pronominal forms as prefixes to both nominal and verbal elements. Both Manoharan and Abbi give due importance to the pronominal prefixes. Manoharan divides the nouns representing body parts into three classes based on the form of the first person singular prefix. The three forms of prefixes that he gives are /ʈɔ-/, /ʈa-/ and /ʈe-/. Similarly there are other pronominal prefixes that Manoharan says works in the same fashion.

Manoharan further finds the personal pronominal system to be in all the three categories with distinction of inclusive/exclusive in the first person. He also observes singular/plural distinction in all the three persons and honorificity in the third person. As all the pronouns have different forms in different cases, as an account for the distribution of these forms, Manoharan gives a classification of verbs 'confined to only the noun morphology of this language'(pp.:75). This classification has been based on the form of the pronominal prefixes that the verbs take i.e. /ʈɔ-/, /ʈe-/, /ʈu-/ and /ʈa-/[1].

While Manoharan's classification based on the distribution of different forms of the personal pronouns is confined only the noun morphology, Abbi's account of the distribution calls for a semantic invocation, as she suggests that 'the distinctions in various forms do not depend upon the simple binary distinction of alienability/inalienability but on various diverse ethno-semantic categories defining the relation between the possessor and the possessed nouns' (Abbi, 2006. pp.:48).

Similarly, Abbi also suggests other 'factors such as part-whole relationship, intimate-non intimate relations, independent household of the possessum, and the notion of possessum being part of the possessor' (ibid).

[1] Manoharan misses the aspiration in all the cases as the first person pronoun is aspirated rendering the transcription as /ʈʰ/ and not just the retroflex plosive.

2.3 Verb Morphology

Based on the imperative form of each of the verbs, Manoharan in his chapter on Verb Morphology classifies Andamanese verbs into 16 classes and adds a list of 'irregular' verbs in the 17th class. The imperative markings of the first 16 classes are as follows:

Verb Class	Imperative Marking	Verb Class	Imperative Marking
1	-be	9	-me
2	-ke	10	-ne
3	-le	11	-ɲe
4	-te	12	-ŋe
5	-ʈe	13	-we
6	-ɖe	14	-ye
7	-ɸe	15	-re
8	-xe	16	-ce/se

Table 7: The Class Marking of Verbs (adapted from Manoharan, 1989 pp.: 92)

Manoharan also notes the existence of copula verbs: /-be ~ bi/, /jiyo/ and /-ɸe ~ -ɸi/. While Manoharan says that '/-be/ being distributed largely, it has been taken as the norm others being allomorphs to it.' Abbi (2003) on the contrary saying that most of the verbs belong to /-k/ class, gives a clearer picture of the verbal morphology with an account of the suffixes marked on the verbs. Great Andamanese distinguishes two tenses i.e. past and non-past. Manoharan again justifies the different forms of the suffixal elements based on the verb classification made by him and does not give any generalized verbs' suffixal schema as suggested by Abbi.

There are several questions that can be raised on the classification and analysis provided by Manoharan. For example, when he suggests that the /-be/ being the general imperative marker and others being only the allomorphs to it, how can speech community have 16 allomorphs to a single morph. Manoharan's data either lacks analysis or something is amiss in his data.

Although Abbi also goes up to saying that the language has six classes (based on the data of the pilot survey in 2001), she has done morphemic analysis to come up to the

verbs' suffixal schema which alone has the potential to account for all the 17 'classes' of Manoharan.

My own analysis confirms Abbi's preliminary schema and more of discussion and explanation to the validation of the schema has been given in the third chapter.

2.4 Syntax

Great Andamanese like other Indian languages also have the most common of the word order i.e. SOV and its implied word order of other grammatical categories. Examples given below will explicate it further.

1	GA	ram konabi ɟiyom						
	MB	ram	MB	ram	MB	ram	MB	
	Gl	Ram	Gl	Ram	Gl	Ram	Gl	
	Egl	Ram is eating a tendu.						
	Src	Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-1						
	Rem	**SIMPLE SENTENCE/ACCUSATIVE**						
2	GA	sita konap^{h}uŋbi ɟiyum						
	MB	sita	MB	sita	MB	sita		
	Gl	Sita	Gl	Sita	Gl	Sita		
	EGl	Sita is eating a ripe tendu.						
	Src	Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-1						
	Rem	**SIMPLE SENTENCE/ADJECTIVE**						

Great Andamanese usually gives sentences which are mostly small clauses and avoid making complex sentences as they do not use conjunctions to add on clauses to get longer and complex sentences. Except for one conjunctive particle /bɔ ~bo/ equivalent to 'and' and 'more', no study of this language has shown existence of any other conjunctions. Thus if a coordinated sentence like 'Lephai will play but not Bie' will be translated as two separate clauses with no conjunction marking in between-

16	GA	alep^{h}ai ereŋk^{h}olebom abie p^{h}o						
	MB	a-lep^{h}ai	MB	a-lep^{h}ai	MB	a-lep^{h}ai	MB	
	Gl	SPEC-Lephai	Gl	SPEC-Lephai	Gl	SPEC-Lephai	Gl	

EGl Lephai will play but not Bie.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-5
Rem **COORDINATION/NEGATIVE**

-thus, showing a gap between the two clauses which proves non-existence of the coordination markings. Even when two nouns are out together, no coordination marking of 'and' is used, rather both the nouns are just juxtaposed to each other, thus-

15 GA aɟoe atoŋ nu təracɔre ulunciko
MB a-ɟoe a-toŋ-nu təracɔr-e eole-inci-k-o
Gl SPEC-Joe SPEC-Tong-PL spring-OBL see-go-CLSM-PST
EGl Joe and Tong went to see the spring.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-5
Rem **COORDINATION**

Even the /bɔ ~bo/ acting sometimes as coordinator is in fact a clause modifier or discourse beginner e.g.:

66 GA bo lɛcik kacil o lɛcit cɔŋ minotəracɛʈʰul
MB bo lɛc-ik ci-l o lɛc-it cɔŋ-ɸ mino-təra-cɛʈʰ-ul
Gl more arrrow-ACC go-PCPL 3SG arrow-ACC find-PST potato-GEN-root-LOC
EGl He went after more of the arrows (and) he found arrow in the root of a potato.
Src Nao Jr./21/01/2006 *The Great Narrative of Phertajido* Narayan A-146
Rem **COORDINATION**

67 GA kʰudi bo o lɛcik tɛrtola eka ʈʰitbolo
MB kʰudi bo lɛc-ik tɛrtol-a eka-ʈʰitbol-o
Gl 3SG and/more arrow-ACC shoot-PST 3SG.CL-search-PST
EGl He searched for more of the shot arrows.
Src Nao Jr./21/01/2006 *The Great Narrative of Phertajido* Narayan A-147
Rem **COORDINATION**

Although no change in SOV order is usually applied for any change in the semantics of the sentence, it is possible in some cases to alter the order of SVO pattern as in the sentence above where the objective noun phrase comes after the verb.

For a good typological overview of the language one can see the table given in Abbi (2006, pp.:85).

Chapter 3
The Verb Morphology

The verb morphology of any language will necessarily cover the topics of tense, aspect and mood (TAM). The TAM markings are suffixed to the verb roots in Great Andamanese. It is the TAM markings that are the most important in the study of Great Andamanese. The second most important thing about the verb morphology is its unique paradigm structure which allows both the subject and object clitic to be prefixed to it. As mentioned in section 1.3.3, Manoharan (1989) gives a list of 17 classes based on the consonantal ending of the verbs in the imperative constructions. Although Manoharan does not give any rationale for dividing the verbs into these classes except for what I could gather from his book and the paper published in 1997. I guess that Manoharan, seeing very appropriately the importance of the pronouns and the 'pronominal prefixes' and the uncertainty of the verbs' TAM markings at several times, was lured by his data to classify the verbs on the basis of its imperative endings. I propose to give an alternative to this analysis; however this alternative is inclusive of an explanation for his data and analysis as well. This alternative also takes on Abbi's (2006) analysis. While agreeing with Manoharan in that the verb shows the classes of consonant, what she calls 'thematic consonant', Abbi (ibid pp.58) says that she could find only six of the classes as against Manoharan's 17 classes. To say that there is no such consonant class will not be fair but to accept the existence of it is also not very clear as is shown in my analysis. Thus, what my findings show below regarding the consonant class or the thematic consonant is only to problematize the topic. However, it finally accepts the notion of a 'thematic consonant' coming as a class marker; it gives only two or at most three such classes.

Manoharan also makes another classification of the verbs on the basis of the pronominal prefixes that the verbs take (Manoharan, 1989 pp.75, Manoharan, 1997). However this analysis he says is "confined to only the noun morphology". Thus he classifies the verbs into four classes based on the pronominal prefixes of the verbs, the pronominal prefixes being /ʈɔ/, /ʈe/, /ʈu/ and /ʈa/[1]. The rationale again is missing. As the verb phrase in Great Andamanese also includes the pronominal prefixes which has been noted here as clitic, a section is devoted to this as well.

3.1. The Concept of Tense, Aspect and Mood

As the tense, aspect and mood are the most important features of any verb morphology; it needs to be discussed first. Tense, aspect and moods are categories on which many linguists have written as it represents a language community's concept of time and how they divide it in relation to their understanding of the world. Based on Reichenbach's theory of time, there are two points in time reference; the reference point (RP) and the event time (ET). The default RP that an event is related to is the time of speech (ST). These relationships are illustrated in following figure, adapted from Givón (2001:286).

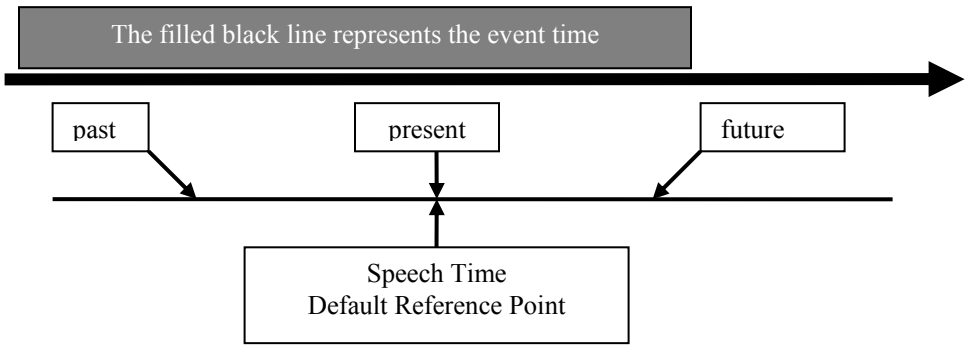


Figure 4: Past, Present and Future: A Tense Division

As is shown above, there are three tense divisions:

- Past: Event time precedes speech time
- Present: Event time is simultaneous with speech time

[1] Manoharan misses the aspiration in his transcription most of the time. The first person singular is denoted by /ʈʰu/, /ʈʰe/, /ʈʰɔ/ and /ʈʰa/.

- Future: Event time follows speech time

An event anchored to current speech act is called an *absolute tense*. The term of relative tense (Givón 2001:286) is used to refer to the cases when tense divisions are anchored to a reference point preceding or following the time of speech. Dahl (1985:29-31) yet another notion, namely temporal frame, in addition to the Speech, Reference and Event notions in order to account for sentences such as:

When I arrived, Peter had tried to phone me twice during the preceding week.

In addition, some scholars (e.g. Bybee et al. 1994) find it difficult to view the so-called present tense as a "tense".

3.1.1. Tense and Aspect in Great Andamanese

3.1.1.1. Tense

All the scholars who have worked on this language agree that GA has a two way time distinction i.e. of past and non-past. While past is marked with a single morpheme of /ɔ/~ /o/ or /kɔ/~/bɔ/, non-past tense is marked with /ɔm/~/om/, /bɔm/~/bom/, and /kɔm/~/kom/.

As is evident from the above, there are variations in the tense marking morphemes. The variation is at two levels. First it is the variation of the vowels, either /o/ or /ɔ/. This variation is found in both intra-speaker speech and inter-speaker speech. That is, the same speaker could give two variations at different times. This variation has nothing to do with the meaning and is only at the phonetic level. But the second variation is something that is very serious that leads to two different analysis of the verb's suffixing phenomena in Great Andamanese. This is what makes Manoharan classify the verbs according to the consonant ending of the verbs in imperative mood, thereby implying that the verbs take a consonant ending before it takes any TAM marking suffixes. As said earlier, Manoharan says there are 17 such 'classes' of the

verbs. Manoharan is silent on whether these verb ending consonants have any meaning attached to it or not. My findings suggest that if there be any verb ending, 'class' marking consonant without any meaning, it is only two or three (/-b/, /-k/ and very slightly, /-l/)[2] and even that is not very clear as is shown below.

3.1.1.1.1. Consonant Class marking, Dialectal Variation or a Hindi Influence?

As said above, there are at least two clear cases of consonant 'class' ending i.e. /-b/ and /-k/.

3.1.1.1.1.1. /-b/ a Consonant 'Class' Marker or a Form of AUX Marking on the Verbs?

Let's first take the case of /-b/. The auxiliary verb in Great Andamanese is /be/ or sometimes its variant /bi/. It would be very easy to say that the verb endings coming after the verb roots and before the TAM markings is a form of this auxiliary. Although this interpretation is very simple and obvious, if one accepts this, the question of /-k/ as a consonant class marker looms large as /ke/ never comes as an auxiliary.

To get to the root of the problem, I started taking a paradigm of verbs (more than thirty verbs were taken) which included constructions of imperative, past and non-past. Besides this there were more verbs taken from sentences of the whole of my data collected (around 115 verbs). The consonant class checking was also done for a list of verbs (more than 200 verbs) that was prepared before I went to the fieldwork. What was found thus can be put like this:

- /-b/ and /-k/ are optional elements in the TAM marking. When these are omitted, the TAM marking is done on the final consonant of the verb root, if any, or there emerges a case of vowel harmony or diphthongization.

[2] I have only two clear cases of the third class in all the syntactic data that I have collected. For the rest of the verbs, the class marking consonant occurring before the TAM marking is optional.

- There are verbs which can take only /-b/ and not /-k/ as a class marking, suggesting that the verb has the /b/ or /k/ in its root.[3]
- /-k/ may be an influence from Hindi verb /karna/ <to do>[4]
- There are cases where both /-b/ and /-k/ are possible, suggesting that /-b/ and /-k/ are dialectal.[5]

Given the findings above, it should be clear that it would be better to treat the language as having no consonant class as Manoharan would suggest. Rather it is either a dialectal variation or simply that /-bom/~/bɔ/ and /-bo/~/bɔ/ are forms of the auxiliary /be/.

3.1.1.1.1.2. /-k/ a Consonant Class Marker, a Hindi Influence or Dialectal Variation?

The /-k/ is either an influence from Hindi verb /karna/ <to do> as they very often mix this in their day to day speech so much so that it has percolated into their own language structure or it is a part of the verb root or it is a dialectal variation.

As most of the dialects of this language are extinct today, it is not possible to say what the right position is in this regard. To take a position will be to be appropriate. So it is better to accept that there is a consonant class marking phenomena in the language, but how does it actually function remains still a question and I will not go much into the details of it as it will be out of the scope of this dissertation its purpose being only to cover the morphology at the descriptive, typological level so that a framework for the verb morphology comes up for a computational model.

---

[3] During our data collection I noticed that my main informant used most of the time /-b/ as class marker while in the later phase he shifted more to using /-k/

[4] As was suggested by two of the informants that /-k/ is spoken in most of the cases by children who get influenced by Hindi as they use Hindi the most in their day to day speech. See MDVI-Gr.4 SI12 Peje saying that /-k/ is used by children. This may also serve as cause of lexicalization of /-k/ as part of the verbs as this also percolated into the speech of the elders. There are several other cases in which the children having been influenced by Hindi influenced their parents' speech. For example /lao/, earlier used for 'demons', has shifted to stand for 'strangers' or simply a generic term for 'human being'.

[5] Nao Jr. suggests in that in some cases /-b/ and /-k/ are dialectal. That is in some cases you use /-b/ and in other /-k/, the variation being dialectal. See MDVI-Gr.3 SI11.

3.1.1.2. Aspect

Aspect, unlike tense, is not concerned with placing events on a time line. Rather, aspect is concerned with making distinctions about the kinds of actions that are described by verbs: progressive actions, punctual actions, habitual actions, etc. Scholars also divide aspects into two types, namely inherent aspect and grammatical aspect (see Dahl, 1985, Givón, 2001). Givón (2001:287) divides verbs into four major categories, depending on their inherent aspectuality, as follows:

- Compact (short duration) verbs: e.g. spit, shoot, jump, hit
- Accomplishment-completion verbs: e.g. arrive/come, obtain/get, die, be born, finish
- Activity-process verbs: e.g. break, bend, step, walk, work, read
- Stative verbs: e.g. be sad, be cold, know, want, be tall, be red

As shown above, inherent aspect can thus be said to encode the typology of states and events in a given language (Saeed 1997:107). The inherent aspectuality closely interacts with the morphological or grammatical aspectuality. It is also often difficult to separate the inherent aspect from contextual influence (Dahl 1985:26f). Grammatical aspect adds a communicative perspective to the events or states described by the inherently aspectual verbs. Aspectuality can thus best be investigated by observing the interaction between inherent and grammatical aspect.

In Great Andamanese it is the inherent aspect that functions and not the grammatical or morphological aspect as there are no other markers on the verbs that could show aspectualities such as say, perfectivity, continuity etc. The aspectual values may also depend very much on the context. As was seen in during our data collection, the

informant rendered the same sentence meaning present continuous, present perfect, future, future perfect etc.[6]

3.1.2. Mood in Great Andamanese

There are at least five distinct morphemes that can be said to constitute five different moods in Great Andamanese. These morphemes are as follows:

Mood	Mood Marking Morpheme
Imperative/Indicative	-e
Prohibitive Negative	-im~em
Conditional	-amo
Stative/Participial	-il~al~el~ul~l
Habitual	-me

Table 8: The Mood Marking Morphemes in Great Andamanese

These mood marking morphemes are all suffixed either directly to the verb roots or to the class marking consonants. Let's take each of the cases separately.

3.1.2.1. Imperative/Indicative

7 GA kʰuro kotrakak ci-be
MB kʰuro kotra-kak ci-b-e
Gl here inside-DIREC come-CLSM-IMP
EGl Come inside.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-3
Rem **IMPERATIVE/DIRECTIONAL**

9 GA ŋaunobe
MB ŋa-uno-b-e
Gl 2SG.SUBJ.CL-sit-CLSM-IMP
EGl Please sit down
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-3

[6] Our data collection method being bilingual, our questionnaires were in Hindi. They also had problem using the standard Hindi as their Hindi was what is called Andamani Hindi and it lacks the rich aspectual system that is used in the standard Hindi.

Rem **IMPERATIVE**

32 GA ʈʰoŋɛ rɔʃe
MB ʈʰo-ŋe-rɔʃe
Gl 1SG.SUBJ.CL-2SG.OBJ.CL-love-IMP(NPST)
EGl I love you.
Src Nao Jr./ 10.01.06/ SI Narayan C-24
Rem **RECIPROCAL CHECK, SIMPLE SENTENCE**

62 GA tekʰamo ʈʰua kʰilil ʈʰɛŋobe
MB tekʰamo ʈʰua kʰilil ʈʰe-eŋo-be
Gl now 1SG here 1SG-come-CLSM-IND
EGl Just now, I came here.
Src Nao Jr./15-11-2005/MD-1, G-12 CIILQ1 T-3/NarayanA-57
Rem **ADVERBIAL CLAUSE**

The imperative and the indicative in Great Andamanese take the same suffix. In fact the imperative mood in Great Andamanese by default is the indicative. While /-e/ is purely for the imperative marking, when it comes with the class marking consonant it may mean indicative as well. The first two examples above show that /be/ is an imperative construction, but the third and the fourth ones show it to be indicative as it expresses a fact. The imperative can be marked with simply /-e/ as well for example:

65 G.A. lɛcbik tɛrtole
MB lɛc-bik tɛrtol-e
Gl arrow-ACC shoot-IMP
EGl Shoot arrow!
Src Peje/22.12.05/MD-6 Gr.5 SI12PEJEVCLSCHK/NarayanC-98
Rem **IMPERATIVE**

61 GA kʰiderbi itborɛ
MB kʰider-bi it-bor-ɛ
Gl coconut-ACC scratch-IMP
EGl Scratch the coconut.
Src Peje/22.12.05/MD-VI-Gr.5-SI12PEJEVCLSCHK/NarayanC-95
Rem **IMPERATIVE**

3.1.2.2. Prohibitive Negative

Prohibitive negative is one of the clearest cases in Great Andamanese. The suffix /-im/ is used to mark these cases. For example:

8 GA ŋu kotrakak ŋucibim
MB ŋu kotra-kak ŋu-ci-b-im
Gl 2SG inside-DIREC 2SG.SUBJ.CL-come-CLSM-PROH.NEG
EGl Do not come in.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-3
Rem **PROHIBITIVE NEGATIVE/DIRECTIONAL**

60 GA ŋu ikjiram
MB ŋu ikjira-m
Gl 2SG say-PROH.NEG
EGl Do not speak.
Src Peje/22.12.05/MD-VI-Gr.5-SI12PEJEVCLSCHK/NarayanC-80
Rem **PROHIBITIVE NEGATIVE**

59 GA ŋut konabi beliŋem
M.B. ŋut kona-bi beliŋ-em
Gl 2SG tendu-ACC cut-PROH.NEG
EGl Do not cut tendu.
Src Peje/22.12.05/MD-VI-Gr.5-SI12PEJEVCLSCHK/NarayanC-75
Rem **PROHIBITIVE NEGATIVE**

Whether prohibitive negative is a Mood or not may theoretically be a question, but by sheer typology of its occurrence as a verbal suffix it stands a good chance of being put under the category of a Mood in Great Andamanese.

3.1.2.3. Conditional

Conditional is marked on the verb roots again in the same way i.e. as a verbal suffix with the morpheme /-amo/, it may have forms like /-bamo/ and /-kamo/ as well, as the class marking consonant may precede it. Thus:

11 GA ŋuʈʰi miʈʰaibi tɛʃe pʰutɛʃamo ʈʰuŋolobom
MB ŋu-ʈʰi miʈʰai-bi tɛʃ-e pʰu-tɛʃ-amo ʈʰo-ŋol-o-b-om
Gl 2SG-1SG.OBJ sweet-ACC give-IMP NEG-give-COND 1SG-cry-EPV-CLSM-NPST
EGl If you do not give me the sweets I will cry.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4
Rem **CONDITIONAL & COORDINATION**

12 GA ŋuʈʰutbatʰebamo ʈʰoŋolobom
MB ŋu-ʈʰut-batʰe-b-amo ʈʰo-ŋol-o-b-om
Gl 2SG-1SG.OBJ.CL-slap-CLSM-COND 1SG.SUBJ.CL-cry-EPV-CLSM-NPST
EGl If you slap me I will cry.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4
Rem **CONDITIONAL & COORDINATION**

13 GA oʈʰobɔyamo ʈʰui tɛrtakom
MB o-ʈʰo-bɔi-amo ʈʰu-i-tɛrta-k-om
Gl 3SG-1SG.OBJ.CL-ask-COND 1SG.SUBJ.CL-3SG.OBJ.CL- tell-CLSM-NPST
EGl If he asks me, I will tell him (the whole story).
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4
Rem **CONDITIONAL & COORDINATION**

38 GA ŋu ɟobi ʃɔrokamo ŋakɛrbin caypʰo
MB ŋu ɟo-be ʃɔro-k-amo ŋa-k-ɛr-beŋ cay-pʰo-ɸ
Gl 2SG sing-IMP sing-CLSM-COND 2SG-cl-GEN-throat bad-NEG-NPST
EGl You sing. If you sing your throat will not be bad.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-30
Rem **CONDITIONAL**

3.1.2.4. Stative/Participial

The verbal suffix /-il/ or its allomorphs as /-al~-el~-ul~-l/ are in fact stative markers. That is its marking suggests that the domain of dynamicity has left the stem and it stands stative. Great Andamanese use this suffix both in verbal classes and in nominal classes. While used with the verbal elements it suggests a sense of participiality or gerunding (with the sense of verb+ing as in English) which in fact is a suggestion of stativity. Examples:

41 GA pʰɔr lobuŋ be ʈʰitumulilyo
MB pʰɔr lobuŋ-be ʈʰi-tumul-il ɟiyo
Gl bamboo long-AUX place-lie-PCPL AUX-EXIST
EGl The long bamboo is lying at the place.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-30

Rem **CLASSIFIERS CHECK, SIMPLE ATTRIBUTIVE SENTENCE**

5 G.A. rizwane konabit beliŋel luk^{h}um beliŋo

MB	rizwan-e	kona-bit	beliŋ-el	luk^{h}um	beliŋ-o
Gl	Rizwan-	tendu-ACC	cut-PCPL	finger	cut-PST

EGl Rizwan cut his finger while cutting the tendu.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-1
Rem **CLAUSE COORDINATION/PARTICIPIALIZATION**

23 GA aboa ameo ikɟirəl ikubeliŋ

MB	a-boa	a-meo	ik-ɟira-l	ik-u-beliŋ-ɸ
Gl	SPEC-Boa	SPEC-Meo	3SG.CL-tell-PCPL	3SG.SUBJ.CL -3SG.OBJ.CL-cut-PST

EGl Boa got it cut by speaking to Meo.
Src Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-15
Rem **PERIPHRASTIC CAUSATIVES, PARTICIPLE**

42 GA akaunol akaunol olamɔ

MB	aka-uno-l	aka-uno-l	o-lam-ɔ
Gl	3SG.CL-sit-PCPL	3SGcl-sit-PCPL	3SG.SUBJ.CL-tire-PST

EGl He got tired while sitting.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-33
Rem **REDUPLICATION CHECK, SIMPLE SENTENCE**

43 GA akaɲal u ʈ^{h}iɟira

MB	aka-ɲa-l	u-ʈ^{h}i-ɟira
Gl	3SG.CL-eat-PCPL	3SG-1SG.OBJ-speak-PST

EGl He told me while eating.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-34
Rem **PARTICIALIZATION, COORDINATION**

3.1.2.5. Habitual

The morpheme /-me/ is another marker that is used as a mood marker in Great Andamanese. It should be noted that Great Andamanese does not allow any passive marking. That is there is no active/passive distinction in Great Andamanese, so far as my data suggests. Instead, in case of the passive construction, they use what has been addressed here as the habitual marker. Examples:

27 GA b^{h}arat^{h}il nu rɛfi iɟume

MB	b^{h}arat^{h}-il	nu	rɛfe-i	iɟi-me
Gl	India-LOC	people	rice-ACC	eat-HABIT

EGl Rice is eaten in India.
Src Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-18
Rem **HABITUAL**

45 GA alepʰai kanʃɔrome eole
MB a-lepʰai kanʃɔro-me eul-e
Gl 3SG-Lephai sing-HABIT see-IMP
EGl See Lephai sing.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-37
Rem **HABITUAL MOOD**

47 GA akarkʰa o rɛfiraʃue ɲyoak enoleme
MB akarkʰa o rɛfi-raʃue ɲyo-ak e-nol-e-me
Gl 3SG.CL-say-PST 3SG rice-cook home-DIREC 3SG.CL-good-EPV-HABIT
EGl He said that cooking at home is good.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-41
Rem **COORDINATION**

3.1.3. Other Verbal Suffixes: Negative Marker

Having thus accounted for what was to be considered as the TAM marking suffixes in Great Andamanese, there remains one morpheme that comes before all these markings i.e. the negative marker. The negative marker in Great Andamanese is /-pʰo ~ pʰɔ/. The negative marker is suffixed immediately after the verb root. All the other suffixes can come only after that. Examples:

6 GA ʈʰu iskule ʈhutconnepʰobe
MB ʈʰu iskul-e ʈhut-conne-pʰo-b-e
Gl 1SG school-OBL 1SG- go-NEG-CLSM-IMP(NPST)
EGl I do not go to school.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-2
Rem **NEGATIVE**

28 GA ebuʈʰu kʰude oʃolopʰo
MB e-buʈʰu kʰude oʃolo-pʰo
Gl 3SG-weak because walk-NEG
EGl He is weak, cannot walk.
Src Nao Jr./ 08.12.05/MD.2 BSL.4/ Narayan C-19
Rem **NEGATIVE/INCAPABILITATIVE**

46 GA rɛfe tunta ɲa-nol acaʃiu ɟupʰo

MB	rɛfe	tun-ta-ɲa-nol	acaʃiu	iɟi-p^{h}o-ɸ
Gl	rice	REFL-TRV-eat-good	whoever	eat-NEG-NPST
EGl	Who does not like to eat good food?			
Src	Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-40			
Rem	**COORDINATION**			

But the negative element can also be added as a prefix to the verb is suffixed for the conditional mood. Thus the following sentence is a perfectly right construct:

11	GA	ŋuʈ^{h}i miʈ^{h}aibi tɛʃe p^{h}utɛʃamo ʈ^{h}uŋolobom				
	MB	ŋu-ʈ^{h}i	miʈ^{h}ai-bi	tɛʃ-e	p^{h}u-tɛʃ-amo	ʈ^{h}o-ŋol-o-b-om
	Gl	2SG-1SG.OBJ	sweet-ACC	give-IMP	NEG-give-COND	1SG-cry-EPV-CLSM-NPST
	EGl	If you do not give me the sweets I will cry.				
	Src	Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4				
	Rem	**CONDITIONAL & COORDINATION**				

3.2. The Verbal Prefixes

As mentioned in the first chapter section1.3.2.4, Great Andamanese is rich in using its pronominal elements as a verbal prefix. Manoharan calls these as pronominal prefixes and divides them into four classes. However, he bases this classification on the basis of only the first person singular pronoun and does not talk anywhere about any other pronominal prefixes. However, Abbi (2006 pp.56) did notice this and suggested that these are pronominal clitics.

3.2.1. Clitics: What is a Clitic?

The case of clitic wherever it has appeared has been a little fuzzy. Several linguists have written on it. The clitics are always confused with affixes. But the case of clitic is always different from affixes. The general features of clitics that can be found in the literature cab be summarized as follows:

Features of a Clitic

- Phonologically bound but syntactically free
- Function at phrase or clause level

- Cannot be integrated into standard discourse without being bound to some other form
- Often have grammatical rather than lexical meaning
- Belong to closed classes like pronouns, prepositions, auxiliary verbs, and conjunctions
- Usually attach to the edges of words, outside of derivational and inflectional affixes
- Often attach to several syntactic categories of words such as head noun, non-head noun, preposition, verb, or adverb
- Phonologically unstressed

A simple definition of a clitic will be to say that it is a 'morpheme that has syntactic characteristics of a word, but shows evidence of being phonologically bound to another word.' Clitics are further divided into two categories i.e. proclitics and enclitics. A proclitic is a clitic that precedes the word to which it is phonologically joined. An enclitic is a clitic that is phonologically joined at the end of a preceding word to form a single unit.

3.2.2. Clitics in Great Andamanese

Great Andamanese uses pronominal clitics in a very rich manner. The pronominal clitics can be used as a proclitic with both nouns and verbs. Here it is the pronominal clitics used as proclitics to verbs that will be considered. Great Andamanese uses both the subject clitic and object clitic as pronominal prefixes to verbs.

3.2.2.1. Subject Clitic

See the following examples:

6 GA ʈʰu iskule ʈhutconnepʰobe
MB ʈʰu iskul-e ʈhut-conne-pʰo-b-e
Gl 1SG school-OBL 1SG- go-NEG-CLSM-IMP(NPST)
EGl I do not go to school.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-2
Rem **NEGATIVE**

8 GA ŋu kotrakak ŋucibim
MB ŋu kotra-kak ŋu-ci-b-im
Gl 2SG inside- 2SG.SUBJ.CL-come-CLSM-

DIREC PROH.NEG
EGl Do not come in.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-3
Rem **PROHIBITIVE NEGATIVE/DIRECTIONAL**

12 GA ŋuʈʰutbatʰebamo ʈʰoŋolobom
MB ŋu-ʈʰut-batʰe-b-amo ʈʰo-ŋol-o-b-om
Gl 2SG-1SG.OBJ.CL-slap-CLSM-COND 1SG.SUBJ.CL-cry-EPV-CLSM-NPST
EGl If you slap me I will cry.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4
Rem **CONDITIONAL & COORDINATION**

29 GA akaːʈa akaɲabom
MB akaːʈa aka-ɲa-b-om
Gl SPEC-girl 3SG.CL-eat-CLSM-NPST
EGl The girl is eating/eats.
Src Nao Jr./ 08.12.05/ MD.2 BSL.5/ Narayan C-20
Rem **SIMPLE NON-PAST SENTENCE**

30 GA ŋu ʈʰɛ ŋɛrenceo
MB ŋu ʈʰɛ ŋ-ɛrenceo
Gl 2SG 3SG 2SG.SUBJ.CL-fight-PST
EGl You fought with me.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-23
Rem **SIMPLE SENTENCE**

In the sentences above Great Andamanese shows the subject clitic attached to the verb root as a prefix. The forms of pronominal subject clitic have phonetic similarity with its nominal form (see chapter 2) in first person and second person while the third person subject clitic may be phonetically different. (See below a list of the subject and object clitic forms.)

3.2.2.2. Object Clitic

30 GA ŋu ʈʰɛ ŋɛrenceo
MB ŋu ʈʰɛ ŋ-ɛrenceo
Gl 2SG 3SG 2SG.SUBJ.CL-fight-PST
EGl You fought with me.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-23
Rem **SIMPLE SENTENCE**

13 GA oʈʰobɔyamo ʈʰui tɛrtakom
MB o-ʈʰo-bɔi-amo ʈʰu-i-tɛrta-k-om
Gl 3SG-1SG.OBJ.CL-ask-COND 1SG.SUBJ.CL-3SG.OBJ.CL- tell-CLSM-NPST
EGl If he asks me, I will tell him (the whole story).
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4
Rem **CONDITIONAL & COORDINATION**

44 GA cyaːk ocikom kɔil tɔ u-ʈʰuncikom
MB cya-k o-ci-k-om kɔil tɔ u-ʈʰu-inci -k-om
Gl what-DIREC 3SG.SUBJ.CL-come-CLSM-NPST later EMPH 3SG.SUBJ.CL-1SG.OBJ.CL-come-CLSM-NPST
EGl Where will he go, later he will come only to me.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-35
Rem **COORDINATION**

The above three examples show the object clitics prefixed to the verbs. Where both subject and object comes together, the object clitic follows the subject clitic.

The clitic sometimes are also dropped in simple sentences. For example:

1 GA ram konabi ɟiyom
MB ram kona-bi iɟi-om
Gl Ram tendu-ACC eat-NPST
EGl Ram is eating a tendu.
Src Nao Jr/ 15.11.05/ MD1 Gr II T3/ Narayan C-1
Rem **SIMPLE SENTENCE**

2 GA sita konapʰuŋbi ɟiyum
MB sita kona-pʰuŋ-bi iɟi-om
Gl Sita tendu-ripe-ACC eat-NPST
EGl Sita is eating a ripe tendu.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-1
Rem **SIMPLE SENTENCE/ADJECTIVE**

5 G.A. rizwane konabit beliŋel lukʰum beliŋo
MB rizwan-e kona-bit beliŋ-el lukʰum beliŋ-o
Gl Rizwan- tendu-ACC cut-PCPL finger cut-PST
EGl Rizwan cut his finger while cutting the tendu.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-1
Rem **CLAUSE COORDINATION/PARTICIPIALIZATION**

In the cases where both the subject and object are pronominals, it is the clitics in most of the cases and not the nominative pronouns that are used e.g.

14 GA ʊʈʰuncikamo ʊʈʰunce
MB u-ʈʰu-n-ci-k-amo u-ʈʰu-n-ce
Gl 3SG.SUBJ.CL-1SG.OBJ.CL-REFL-come-CLSM-COND 3SG.SUBJ.CL.-1SG.OBJ.CL-REFL-come-IMP
EGl If he can come, he should/must come.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4
Rem **CONDITIONAL & COORDINATION**

39 GA ŋoiʃɔro
MB ŋo- i-ʃɔr-o
Gl 2SG.SUBJ.CL-3SG.OBJ.CL-sing-PST
EGl You sing.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-30
Rem **IMPERATIVE**

36 GA ŋaunobe
MB ŋa-uno-be
Gl 2SG-sit-IMP
EGl Sit.
Src Nao Jr./ 10.01.06/MD.1 Gr.16 T.1-4/ Narayan C-29
Rem **SIMPLE SENTENCE/IMPERATIVE**

The following table gives a list of the forms that are used as subject and object clitics:

Person	Number	Excl/Incl	Subject Clitic	Object Clitic
1st	Singular	Exclusive	ʈʰu, ʈho, ʈʰa, ʈʰe, ʈʰ, ʈʰut	ʈʰu, ʈʰa, ʈʰɛ, ʈhi
		Inclusive	ʈʰɛŋut	
	Plural	Exclusive	me, mut	me, mut
		Inclusive	meŋut	
2nd	Singular		ŋu, ŋa, ŋe, ŋɛ, ŋi, ŋ, ŋut	ŋu, ŋa, ŋe, ŋɛ, ŋi, ŋut
	Plural		ŋale, ŋole	
3rd	Singular		u, o, a, e, dut, uku, aka	aka, ɛk, ek, it, ut, et
	Plural		dunot, nu	na

Table 9: List of Great Andamanese Subject and Object Clitic Forms

Note: The vacant forms could not be found in the data resources. They may or may not be present in the language.

3.2.3. **Causatives**

The causative or the valence increasing morpheme (sometimes also called as transitivizer) is another morpheme that comes before a verb and increases its argument valancy. For example:

24	GA	araːmliʃu aranʃulut^{h}u tabeno			
	MB	araːmliʃu	aranʃulut^{h}u	ta-beno-ɸ	
	Gl	sister	younger brother	TRV-sleep-NPST	
	EGl	The elder sister makes her younger brother sleep.			
	Src	Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-16			
	Rem	**CAUSATIVES**			
25	GA	at^{h}ire teobi ata emp^{h}ilo			
	MB	a-t^{h}ire	teo-bi	a-ta	emp^{h}il-o
	Gl	SPEC-child	crocodile-ACC	3SG.OBJ.CL-INSTR	kill-PST
	EGl	The child got killed by the crocodile.			
	Src	Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-17			
	Rem	**CAUSATIVE/INSTRUMENTAL/ACCUSATIVE**			
22	GA	a-kaba a-buro girəl a-bie ta-eluk^{h}-o				
	MB	a-kaba	a-buro	ik-ɟira-l	a-bie	ta-eluk^{h}-o
	Gl	SPEC-Kaba	SPEC-Buro	3SGcl-tell-PCPL	SPEC-Bie	TRV-pick-PST
	EGl	Kaba got bie picked up by asking buro				
	Src	Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-15				
	Rem	**PERIPHRASTIC CAUSATIVES, PARTICIPLE**				

As seen in the above examples, the causative marker or the transitivizer comes before the verb root. But when both the clitic and the transitivizer may either precede or follow the clitics, e.g.

21	GA	amimitun at^{h}ire ta-əkaŋa			
	MB	a-mimi-tun	a-t^{h}ire	ta-aka-ŋa-ɸ	
	Gl	SPEC-mother-REFL	SPEC-child	TRV-3SG.OBJ.CL-VI-eat-PST	
	EGl	The mother fed her child			
	Src	Nao Jr/ 08.12.05/ MD.2 BSL.3/ Narayan C-14			

Rem **CAUSATIVES/PASSIVES/INCAPABILITATIVES**

25 GA at^{h}ire teobi ata emp^{h}ilo

MB	a-t^{h}ire	teo-bi	a-ta	emp^{h}il-o
Gl	SPEC-child	crocodile-ACC	3SG.OBJ.CL-INSTR	kill-PST

EGl The child got killed by the crocodile.
Src Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-17
Rem **CAUSATIVE/INSTRUMENTAL/ACCUSATIVE**

But in most of the cases the transitivizer /ta-/ precedes the verb root.

3.2.4. **Reflexives**

The reflexive morpheme /tun~tum~un~um~em/ in Great Andamanese is also prefixed to the verb root as is exemplified below:

26 GA beibi tunʈɔlo

MB	bei-bi	tun-ʈɔl-o
Gl	bottle-ACC	REFL-break-PST

EGl The bottle broke.
Src Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-17
Rem **REFLEXIVE/INCHOATIVE**

It follows the pronominal clitic whenever it occurs in the verb phrase e.g.

37 GA caːy k^{h}udi ŋutunceiko

MB	caːy	k^{h}udi	ŋu-tun-cei-k-o
Gl	what	for	2SG.SUBJ.CL-REFL-angry-CLSM-NPST

EGl Why are you becoming angry?
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-29
Rem **INTERROGATIVE**

But the reflexive precedes the transitivizer element if both come together in the verb phrase-

46 GA rɛfe tunta ɲa-nol acaʃiu ɟup^{h}o

MB	rɛfe	tun-ta-ɲa-nol	acaʃiu	iɟi-p^{h}o-ɸ
Gl	rice	REFL-TRV-eat-good	whoever	eat-NEG-NPST

EGl Who does not like to eat good food?
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-40
Rem **COORDINATION**

3.2.5. **Negatives**

Although the negative particle /pʰo~ pʰɔ/ is suffixal in the verb phrase, but in some cases, such as a conditional, it may also precede the verb root e.g.

11	GA	ŋuʈʰi	miʈʰaibi	tɛʃe	pʰutɛʃamo	ʈʰuŋolobom
	MB	ŋu-ʈʰi	miʈʰai-bi	tɛʃ-e	pʰu-tɛʃ-amo	ʈʰo-ŋol-o-b-om
	Gl	2SG-1SG.OBJ	sweet-ACC	give-IMP	NEG-give-COND	1SG-cry-EPV-CLSM-NPST
	EGl	If you do not give me the sweets I will cry.				
	Src	Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4				
	Rem	**CONDITIONAL & COORDINATION**				

3.3. **The Existential Auxiliary Verb**

The auxiliary verb in any language most of the time make a separate case in itself. It is true of Great Andamanese also. There are two morphemes for showing existential cases in Great Andamanese through auxiliary verb <to be>, namely /be~bi/ and /jiyo~jio/. While the first form can be inflected for TAM markings, the second one is static and can be used in both past and non-past tenses rendering an English equivalent of <is/am/are> or <was/were>.

3.3.1. **The Auxiliary Verb /be/**

The auxiliary verb /be~bi/ can stand alone to refer to an existential statement e.g.

17	GA	kʰitani	ŋiyo	be
	MB	kʰitani	ŋiyo	be
	Gl	how	2SG.EXIST	AUX
	EGl	How are you?		
	Src	Nao Jr./ 08.12.05/ MD.2 BSL.3/ Narayan C-8		
	Rem	**WH-QUESTION**		

18	GA	di	aʃyu	bi
	MB	di	aʃyu	bi
	Gl	3SG	who	AUX
	EGl	Who is he?		
	Src	Nao Jr./ 08.12.05/ MD.2 BSL.3/ Narayan C-9		
	Rem	**YES-NO QUESTION**		

20 GA a-tʰire eʈɔlɔʈmobe
MB a-tʰire eʈɔlɔʈmo be
Gl SPEC-child fair AUX
EGl The child is fair.
Src Nao Jr./ 08.12.05/ MD.2 BSL.3/ Narayan C-13
Rem **SIMPLE ATTRIBUTIVE SENTENCE**

34 GA baksa tuttaraːl ʈʰutɟulu-be
MB baksa tuttaraːl ʈʰut-julu be
Gl box LOC 1SG.GEN-cloth AUX
EGl My clothes are on the box.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-25
Rem **LOCATIVE**

This auxiliary can also take inflectional TAM marking suffixes as well as verbal prefixes e.g.

63 GA u bazaːre akabilbe
MB u bazaːr-e aka-bi-l-b-e
Gl 3SG market-LOC 3SG.SUBJ.CL-be-PCPL-CLSM-IND
EGl He is always in the market.
Src Nao Jr./15-11-2005/MD-1, G-12 CIILQ1 T-3/NarayanA-60
Rem **SIMPLE EXISTENTIAL SENTENCE**

3.3.2. **The Auxiliary Verb /ɟiyo/**

The auxiliary verb also /ɟiyo/ comes as an existential auxiliary verb but it never comes in an inflected form. However, it may take another auxiliary verb /be/ after it. Examples:

33 GA kitaːb ʈʰica tuttaraːl ɟiyo
MB kitaːb ʈʰica tuttaraːl ɟiyo
Gl book table LOC AUX
EGl Book is on the table.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-25
Rem **LOCATIVE**

35 GA kitab baksa tumbol ɟiyo
MB kitaːb baksa tumbol ɟiyo
Gl book box under AUX
EGl The book is under the box.

Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-26
Rem **LOCATIVE**

40 GA pʰɔr lobuŋ be kʰulol ɟiyo

MB	pʰɔr	lobuŋ	be	kʰulol	ɟiyo
Gl	bamboo	long	AUX	there	AUX-EXIST

EGl Bamboo is long. (It) is there.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-30
Rem **CLASSIFIERS CHECK, SIMPLE ATTRIBUTIVE CLAUSE COORDINATION**

3.4. Defining the Verb Phrase in Great Andamanese

Having covered both the suffixes and prefixes concatenated to the verb root, most of the part of the verb morphology has been covered. Now there is a need for definition of a verb phrase so that the term of verb phrase in Great Andamanese can be determined. The verb phrase in Great Andamanese will constitute of its suffixes which includes the statement negative particle, the 'consonant class' class marking and the TAM markings. As its prefixal elements it will have the clitics, the transitivizer morpheme, the reflexives and in a particular case, a statement negative particle. The structured that we have reached thus far shows that Great Andamanese is more of an agglutinative language. A paradigm structure of the verb morphology has been given below.

3.5. A Paradigm for the Verb Phrase in Great Andamanese

The conclusion that we reached thus gives us the structure of the verb paradigm as illustrated below:

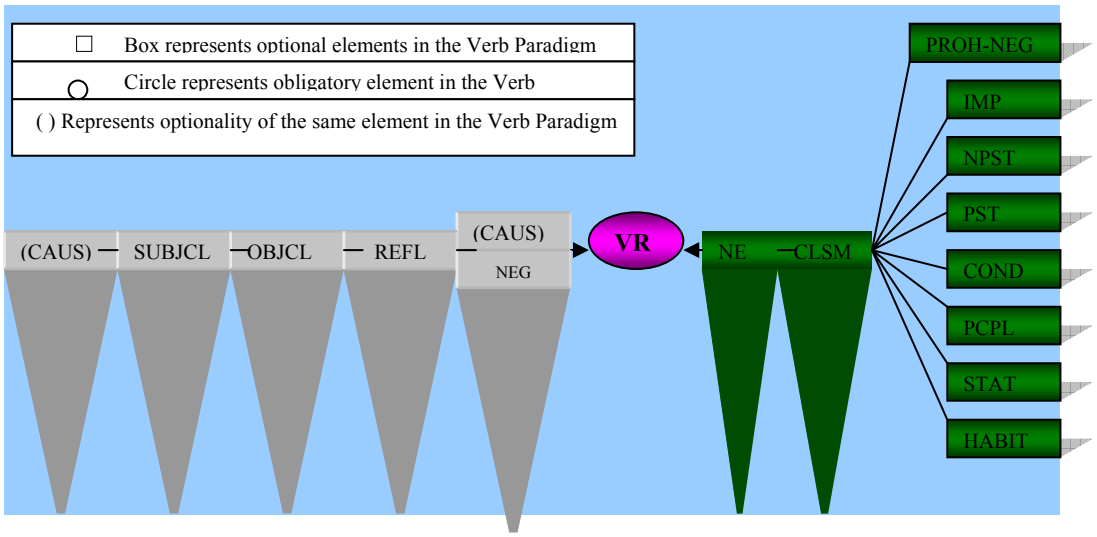


Figure 5: Verb Paradigm Schema of Present Great Andamanese

The verb paradigm as shown above in the diagram shows that a verb phrase has obligatorily to have at least one verb root while it can have other elements optionally marked on it. Thus a Verb Root (VR) can have a maximum of four prefixes and three suffixes. The prefixes of the verbs are the causative, subject clitic and object clitic in the same order followed by reflexive and causative morphemes respectively. The negative particle is conditional and may or may not replace the causative immediately preceding the verb root. The order of the suffixes on verb is as again as given in the figure above i.e. if there is any statement negative it will come just after the verb root followed by class marking and any of the eight TAM marking morphemes.

A list of all the prefixes and suffixes has been provided in Appendix III.

Chapter 4

The Computational Framework

4.1 The Model of Great Andamanese Verb Analyzer

Great Andamanese Verb Analyzer (GAVA) has five modules to undergo processing before it gives any output. These are:

- Pre-Processor
- Verb Root Check
- Prefix Check
- Suffix Check
- G-Filters

However, like the verb paradigm, not all the modules are obligatory and if the proper result is found just after the first or the second module, the output may be reached and other modules may be skipped. This way processing is fast and economical. If the correct output is not found after the processing of the first module, the output of this module goes as input to the next module. Thus the output of one module is taken as input of next module. The interaction and collaboration of these modules are not limited only to this. The interaction is deeper as sometimes the processing of one element is repeated twice or thrice, through each of the modules till it gets the final result.

4.1.1 Model Diagram

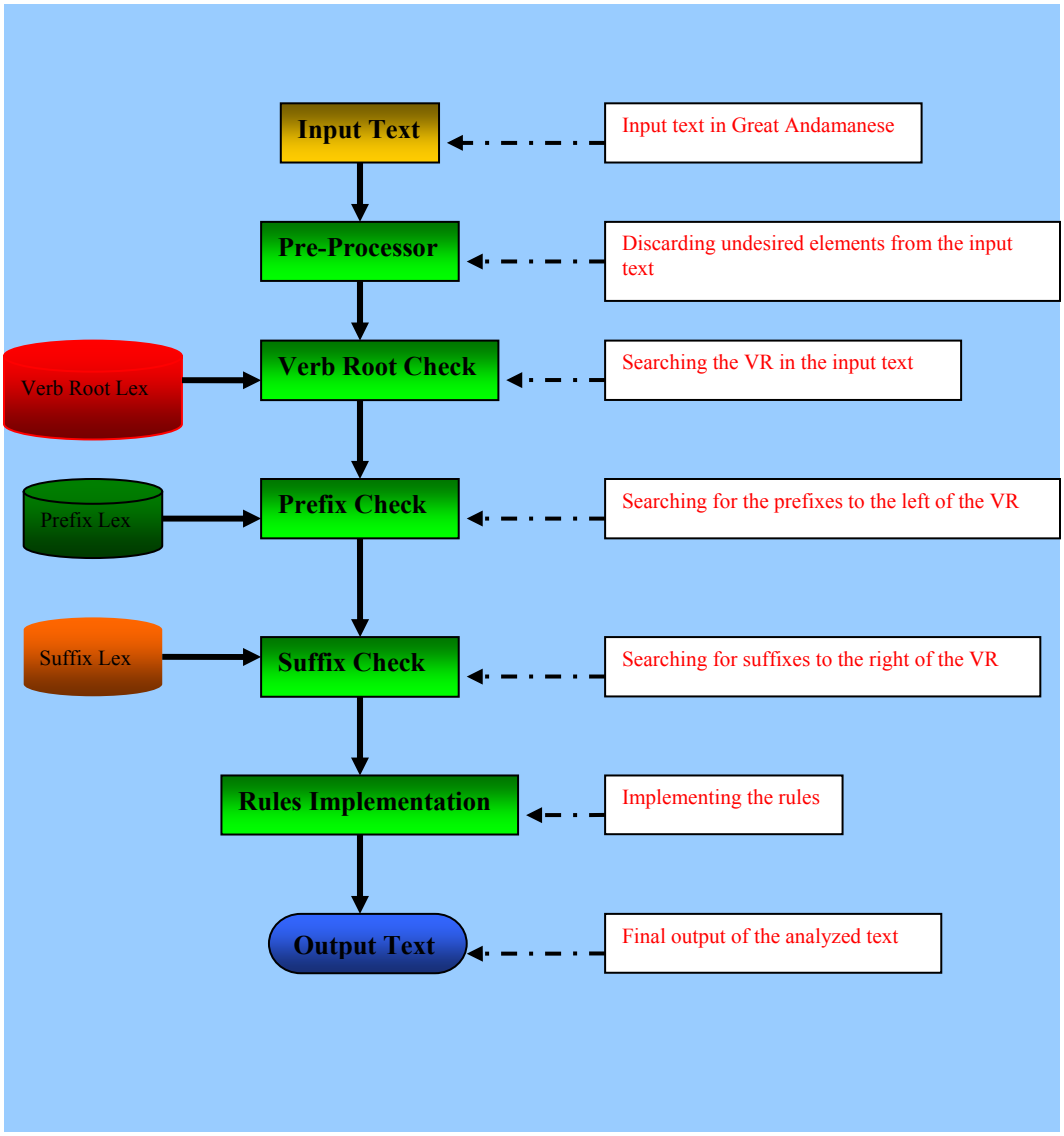


Figure 6: A Model Diagram of the Great Andamanese Verb Analyzer

4.1.2 Brief description of the processes

The input goes through five processes before it gets the final result. As illustrated in the diagram in the previous page, these five processes are in fact five modules that an input undergoes.

The pre-processor module first filters the input and checks whether any unwanted elements are not in the input text. If this is the case, it either corrects the input or leaves it as it is for the consideration of user.

The verb root module searches for the verb roots in the input text and segments them from the string. The remaining part of the input string is sent for further analysis in the next modules.

The prefix module takes the elements that are to the left of the verb root and analyzes them by matching each of the possible strings with the prefixes in the prefix lexicon and stores the results for display.

The suffix module takes the elements that are to the right of the verb root and analyzes them matching each of the possible strings with the suffixes in the suffix lexicon and stores the results for display.

The G-Filters module is the last module that implements the grammatical rules. If the system has not found the right analysis of the input text or there is some ambiguity or violation of some rules, these are checked through rules here.

The final result is displayed as Unicode HTML on a JSP web front.

4.2 The Great Andamanese Verb Paradigm

The verb paradigm of Great Andamanese was drawn in the last chapter through an empirical and evidential method. For a reference, it has been reproduced here:

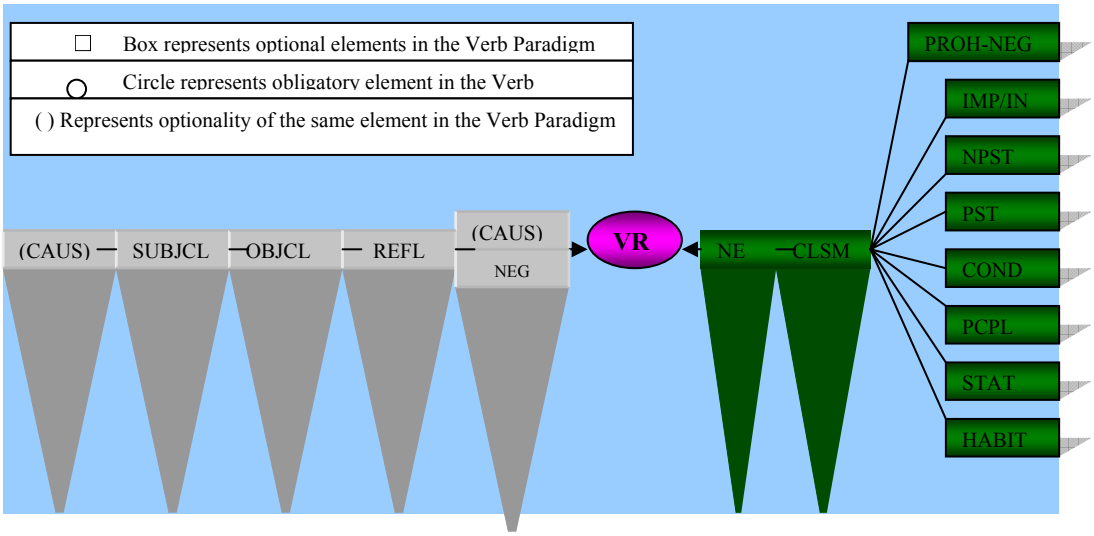


Figure 5: Verb Paradigm Schema of Present Great Andamanese (*reproduced*)

4.2.1 Dynamics of GA Verb Forms

As is shown in the figure above, the verb paradigm has a linear structure with the centrality of verb root (VR, indicated with a violet oval). The verb root is the only element that is obligatory in a verb phrase and all the other elements are optional. There are some rules that govern the positioning of the affixes given below:

4.2.1.1 Suffixes

There are three layers of suffixes attached to a verb root. These three layers are as follows:

NEG→CLSM→TAM

The first and the second layers have only two morphemes while the third one has eight morphemes. The morphemicity of the second element, however, is doubtful as this researcher has not yet reached any conclusion about its semantic content (For more on class markers, see chapter 3, Section 3.1.1.1.1).

In this sequence of the suffixes to the verb root only the negative mood stands out from the usual position of the TAM markings. TAM markings usually come at the end of the verb root. The class marker, coming in between the TAM markings and negative marking, works as a bridge between the two elements. However it may look curious to find the negative mood only coming at this position in the verb paradigm, it has some practical reasons if one looks at the language structure as it is. It may also serve as evidence against the existence of class markers. But here, as the researcher is concerned mainly about the structure of a paradigm, the dissertation will concentrate only on the structure and positioning of the suffixes.

The sequence means that the suffixes come in this order only and not in any other order. However, given the optionality of all the suffixes, one or the other may be dropped, without disturbing the sequence. Thus, we may have examples like:

VR+NEG+CLSM+IMP/IND/PST/NPST/PCPL/COND/STAT/HABIT

Or

VR+CLSM+ PROH.NEG/IMP/IND/PST/NPST/PCPL/COND/STAT/HABIT

Or

VR+ PROH.NEG/IMP/IND/PST/NPST/PCPL/COND/STAT/HABIT

Or

VR

It is obvious that two negative particles cannot come simultaneously in the same verb phrase, thus nullifying the possibility of NEG and PROH.NEG coming together in the same VP.

Looking at the structure statistically, thus we have at most three suffixes coming after the verb root. This draw is important as it is useful in constraining the computational application so that the efficiency and speed of the program remains at the edge.

4.2.1.2 Prefixes

As shown in the illustration above, there are five linear layers of prefixes preceding the verb root:

CAUS → SUBJ.CL → OBJ.CL → REFL → CAUS/NEG

Again there is the optionality in all the prefixes. The causative prefix has two options in terms of its position in the verb phrase. It can take either of the two positions. However, it cannot occupy both the places. Thus in the prefixes, one can have a maximum number of prefixes i.e. five in the following order:

CAUS → SUBJ.CL → OBJ.CL → REFL→ NEG→ **VR**

However, if this order of prefixes occurs in the verb phrase, the negative and prohibitive negative suffixes must be dropped from the verb phrases as they will be redundant and a maximum of only two suffixes will be allowed. Thus **a verb phrase with a maximum number of suffixes and prefixes** will have the following kind of structure:

CAUS → SUBJ.CL → OBJ.CL → REFL→ NEG→ **VR** ← CLSM ← TAM

Other orders of prefixes in the verb phrase can have structures like the following:

CAUS → SUBJ.CL → OBJ.CL → REFL→ NEG→ **VR**

CAUS → SUBJ.CL → OBJ.CL → REFL→ **VR**

CAUS → SUBJ.CL → OBJ.CL → **VR**

CAUS → SUBJ.CL →**VR**

CAUS → **VR**

VR

4.2.2 POS Tags for Great Andamanese Verbs

All the verbs have been tagged with its meaning and an additional identifier of VR in the lexicon. No classification of the verbs as per transitive/intransitive or on any other criteria has been made. However, there is a case of consonant class marking of verbs as said by Manoharan (17 classes) and Abbi (*see previous chapter*), no such criteria has been followed here for reasons given in the previous chapter.

4.3 Linguistic Resources

The linguistic resources have been prepared all on the basis of first hand data collected by the researcher and compared with others available. The program uses lexicons that are basically text files of small sizes.

4.3.1 Tagged Lexicons

There are three lexicons for the use of the Great Andamanese Verb Analyzer.

- Verb Roots
- Prefixes
- Suffixes

All these lexicons are tagged properly as illustrated in the data used in the previous chapters (See *Abbreviations* list). The prefixes and suffixes have an additional tag of PREF and SUFF respectively. This is for specific use of the program and is also displayed in the output text.

The verb root lexicon contains about 120 verb roots[1]. The longest verb root in Great Andamanese so far found is of three syllables and containing eight characters. The frequency of monosyllabic roots is more than disyllabic roots and there is much less tri-syllabic roots. All these roots have been arranged in the lexicon in an ascending

[1] The research in this language is still on. I have taken these verb roots after a painstaking consideration of its forms in several sentence types. Still there may be some discrepancies as one may always differ from a verb form's root representation. There may be more root forms added to it as the analysis for more verbs and sufficient data sources are available.

order of the number of characters present in the root to facilitate better search by the program.

There are a total of 52 prefixes and 20 suffixes at present. The number of affixes has grown up because there are allomorphic variations. Thus a morpheme with a gloss of 1SG.SUBJ.CL has 6 variations, 1SG.OBJ.CL has 4 variations etc. The following table gives a list of allomorphic variations in the verbal clitics.

Name of the Prefix	No.of Variants	Variant Forms
1SG.SUBJ.CL	6	ʈʰu, ʈʰa, ʈʰo, ʈʰe, ʈʰ, ʈʰut
1PL.SUBJ.CL	2	me, mut
1SG.OBJ.CL	4	ʈʰu, ʈʰa, **ʈʰi**, ʈʰɛ,
2SG.SUBJ.CL	7	ŋu, ŋa, ŋe, ŋɛ, ŋi, ŋ, ŋut
2SG.OBJ.CL	4	ŋu, ŋa, ŋe, ŋɛ, ŋi, ŋut
3SG.SUBJ.CL	7	**u**, **o**, **a**, **e**, aka, uku, dut
3SG.OBJ.CL	8	aka, **ek**, **ɛk**, **ek**, **ik**, **it**, **ut**, **et**, **i**
3PL.SUBJ.CL	2	nu, n
3PL.OBJ.CL	NA	Not Available

Table10: A list of pronominal verbal clitics in Great Andamanese[2]

(*The entries in* **bold** *characters show single occurrence of the clitic as against dual entries of others which are the source of ambiguity in a finite state automaton*.)

4.3.2 Rules

At present there is only one rule implemented. This rule takes care of the ordering problem of the prefixes that emerges due to dual entries of the same

[2] The list is not final as it is based on a limited source of data. There may be more or less variants, their names and forms. More specific study on this topic is warranted.

prefix, namely the clitics. For example, /ʈʰu/ can be used as both subject clitic and as object clitic. Similarly there are other clitics that have dual entries. This problem can be solved by the constraint of ordering. In a verb phrase there can be no more than two clitics. These two clitics has necessarily to be different and in the order of subject and object clitic. That is the first clitic has necessarily to be subject and the next object. This solves the problem of the ambiguity emerging out of the dual tags of the same entry.

If a single clitic in a verb phrase is found, it is assigned the tag of subject clitic. However, it is not always necessary that the single clitic in the VP is subject clitic. It can be an object clitic as well, if the subject is omitted or is not a pronominal category in the verb phrase. In this case, the solution lies only in the context of the whole clause. As the present endeavor is does aspire to do contextual analysis, it is left for future research and development.

There are also morphophonemics involved in the verb morphology of Great Andamanese, however small in number. But these are also left for the future, given the limited time and resources available to the researcher.

4.4 Implementation Strategies

The program has been prepared on a Windows platform with tools and techniques as described below. This program however is platform independent and can run on any platform.

4.4.1 An Overview of the Tools and Techniques Used

The following is an overview of the tools and techniques used in developing the program.

- Front end
    - JSP, HTML, CSS, Java Script

- Java Objects
    - Pre-processor
    - Analyzer
        - Search Parts()
            - Verb root
            - Prefixes
            - Suffixes
        - gFilter()
        - reorder()
- Back-end
    - Data files stored in UTF-8
- Webserver
    - Apache-Tomcat

4.5 Front End

At the front end of the program the technologies like the JSP, HTML, CSS, Java Script have been used. The following is a brief intro to these technologies and how they have been implemented in developing the program.

The front end opens in a web browser that is based locally on the user's computer. The URL opens the JSP file located on the host computer usually at the path <C:\Program Files\Apache Tomcat 4.0\webapps\andamanese>. The browser, with the help of the java-webserver, reads the *andverbs.jsp* file. To understand the structure of the *andverbs.jsp* file, we need to have a look at the following.

4.5.1 Java Server Pages (JSP)

Java server pages are html pages which uses Java objects embedded in the html code. The java server page used here utilizes all of the four items discussed above. It uses

first, the html coding convention and initializes the style sheet, the java objects from AVT*agger.class* as servlets.

4.5.1.1 Java

Java is an object-oriented programming language developed initially by James Gosling and colleagues at Sun Microsystems. The language, initially called Oak (named after the oak trees outside Gosling's office), was intended to replace C++, although the feature set better resembles that of object C. Java should not be confused with JavaScript, which shares only the name and a similar C-like syntax. Sun Microsystems currently maintains and updates Java regularly.

Java is a network-oriented programming language that is specifically designed for writing programs that can be safely downloaded to your computer through the internet and immediately run without fear of viruses or other harm to your computer or files. Using small Java programs (called "Applets"), web pages can include functions such as animation's, calculators, and other fancy tricks. Java programs are of three kinds –

- Stand-alone executable programs
- Applets
- Servlets

4.5.1.2 Applets

An applet is a small program written in Java and embedded in a HTML page using the tag <APPLET>. It safely downloads on the client machine and runs using the client Java runtime environment. Applets differ from full-fledged Java applications in that they are not allowed to access certain resources on the local computer, such as files and serial devices (modems, printers, etc.), and are prohibited from communicating with most other computers across a network. The current rule is that an applet can only make an internet connection to the computer from which the applet was sent.

4.5.1.3 Servlets

A servlet is an application program, written in Java and executed on a java compatible web server. A reference to a servlet appears in the markup for a web page, in the same way that a reference to a graphics file appears. The web-server executes the servlet and sends the results of the execution (if there are any) to the web browser as HTML text.. Servlets act as go-between between the clients and the servers using a lot of threading and buffering techniques. These are faster, safer and considered critical for high performance secure websites.

4.5.1.4 Cascading Style Sheets

CSS, short for Cascading Style Sheets, a new feature being added to HTML that gives both Web site developers and users more control over how pages are displayed. With CSS, designers and users can create style sheets that define how different elements, such as fonts, spacing, aural cues, headers and links, appear. These Style Sheets can then be applied to any Web page.

Here are the CSS has been used to bring text in a particular font namely the Lucida Sans Unicode. Another font named Arial Unicode MS can also be used for the purpose of entering the input text in Great Andamanese.

4.5.1.5 HTML

HTML or Hyper Text Mark-up Language, is the base of the front end of the interface on which other objects namely that of Java Objects and CSS has been embedded. HTML is the oldest language that served the purpose of developing web pages.

4.5.2 Java Objects

The JSP file called the *andverbs.jsp* uses a java object called AVTagger which uses the services of Pre-processor. These *.class* files are the compiled programs written in Java that analyze the input text in Great Andamanese. The *Pre-processor* object first

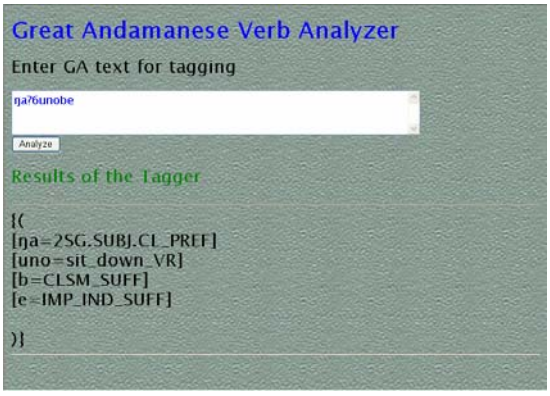
filters the input text and checks whether the input text is a potential Great Andamanese text or not. The *AVTagger* object is in fact the analyzer program that processes the input text after it has been filtered by the pre-processor. There are several modules in this program that take care of several processes undergoing during the mechanical analysis.

As described briefly above, there are five modules of the GAVA program. Below is given the description of each of the modules. Wherever necessary, important snippets of the code written in Java have been reproduced here within <>.

4.5.2.1 Pre-processor

As is illustrated in the block diagram above, the input first goes to the pre-processor module and checks whether any undesired elements such as punctuation marks or other control characters, numbers etc are not given in input. If this is the case, either it corrects the input text itself or removes them from going into further analysis. Thus say for example, if someone has given an input like /ŋaʔɓunobe/ which has unwanted characters /ʔɓ/, the system will filter these out and send the rest of the string of characters for processing in the next module.

The pre-processor prepared has a separate entity in itself and is used by the analyzer code as an object. The result of this pre-processor is given as output to the analyzer program and the analyzer program processes it further morphologically.



Screen Shot 1: Functioning of the Pre-processor

However, in the code written for the program i.e. the *AVTagger*, this has been handled on a default basis. That is, the system recognizes only those strings and

characters that are supported by its lexicon and filters out others. Thus, the declaration for search of input word in the verb root lexicon is made as shown below -

```
<
        inputWord = search_Parts(inputWord, "verbroot", VERBROOTS);
>
```

The input word is searched in the verb root lexicon and if a verb root is found, it ignores the undesired characters and only processes those strings that are available in its lexicon. By the same declaration, the verb roots are also found, and if the verb root is not found, the input is returned as it is, without going into any further analysis.

4.5.2.2 Analyzer

The input thus comes filtered from the pre-processor module to be analyzed by another java object called *AVTagger*. For convenience, this can be called the analyzer. It is this file that is the most important to the program. Two Java APIs from the Java library have been imported to be used in this object.

The analyzer uses several functions and methods to analyze the GA verb (which can have a maximum of 9 elements – 5 prefixes, 3 suffixes and 1 verb root)

4.5.2.3 parseVerbs

This is the main calling function which gets all the work done by using services of other functions/methods. This function first gets the preprocessing done on the entire text. It then tokenizes the output of the pre-processor based on space character (the assumption is that the text can have multiple words separated by space). Then by calling the search_Parts() function, it processes each word for verb, and affixes (to a maximum of 5 prefixes and 3 suffixes).

4.5.2.3.1 Search Parts

The analyzer analyzes each of the entries in parts. As an input may contain several categories of linguistic information that are distributed at different locations in the input, they are divided into parts. The search is then for the parts starting from the

whole of the input to the last available string in the input text until the search is complete or there are no characters left to be searched and matched with the lexicon (or first five prefixes and first three suffixes have been searched). The search is processed in three modules as described below. The search_Parts module assumes the role of searching verbs, prefixes or suffixes when an appropriate call is made for each kind of search.

4.5.2.3.1.1 Verb Roots

The direct call to search_Parts() for searching verbs is as follows –

```
inputWord = search_Parts(inputWord, "verbroot", VERBROOTS);
```

The input thus coming filtered from the pre-processor module to the verb root check module. Here a string of characters divided by space is taken for consideration. This module looks for a verb root in the search string. It looks in the verb root lexicon and searches the existence of the verb roots, matching the input string with the verb roots stored in the lexicon, one after another. If the verb root is found in the lexicon and if there are characters still left to the right or left of the verb root, it is prompted for further analysis in the next module.

As the verb root can be at any position in the verb phrase, the search is tuned for all the positions i.e. in the middle or to the right or the left edge of the verb phrase. The position also decides the occurrence of the prefix or suffix in the verb phrase. That is if the verb root is in the middle of the verb phrase, it is certain that the characters to the left of it are prefixes and the characters to the right of it are suffixes. That is the decision of the occurrence of the characters that are potential prefixes and suffixes are done at this module itself.

If there are no matches of verb root found in the search string, the module takes it as a non-verbal element and returns the input string as it is, without any analysis. In this

case, further processing is stopped and the next modules are skipped to return the result as it is.

Code Snippet:

```
if ( inputWord.length()>0){
                //If the length of input word is more than 0
                //if not a VR then search for affixes to the right and left
                //if the verb was found in the middle, prefix and suffix both are present

        if ( inputWord.indexOf("_")>-1){
                //finding the key in the lexicon after the underscore '_'
        post=inputWord.substring(inputWord.indexOf("_")+1,inputWord.length());
                //Defining the suffixal string
        pre = inputWord.substring(0,inputWord.indexOf("_"));
                //Defining the prefixal string
                errmsg = errmsg +"pre=" +pre+ "   post="+post+"<br>";
                //Defining the error message for display
                search_affix("suffix", post, SUFFIXES);
                        //initializing the search function on the suffixes
                search_affix("prefix", pre, PREFIXES);
                        //initializing the search function on the prefixes
                }
        }
```

4.5.2.3.1.2 Prefixes

The indirect call to search_Parts() for searching prefixes is as follows –

```
search_affix("prefix", inputWord, PREFIXES);
```

The verb root check module prompts what is left to the left of the verb root to the prefix module. If there are no characters found to the left of the verb root this module will be skipped. As mentioned in the verb paradigm above, the dynamics of the Great Andamanese verbs allows at most five prefixes in a specific order. The order cannot

be violated except for what has been described above. If the system recognizes more than five prefixes, it will either give a wrong output or miss out the extra prefixes. After it has found all the prefixes, it stores them in the buffer for the G-Filter to check whether they require any change according to the rules specified.

Code Snippet

This comes forward from the code snippet given above for the verb root search. Thus just the else function for the statement made for the verb root search can suffice the need for the prefix and suffix search. For example, the following code searches for suffixes-

```
else if (searchType.equals("prefix")){
//initializing the else function on the verb root search statement.

search_affix("prefix", inputWord, PREFIXES);
//searching the prefixes on the prefix lexicon
```

4.5.2.3.1.3 Suffixes

The indirect call to search_Parts() for searching suffixes is as follows –

```
search_affix("suffix", inputWord, SUFFIXES);
```

The input to this module also comes from the second module of the verb root check. The characters to the right of the verb root are taken for consideration in the suffix module. There can be at most three suffixes after the verb root. This number is taken as constraint. The system will recognize only three suffixes after the verb root. The GAVA here expects that the input is correct and analyzes up to three suffixes. If more

than three suffixes are given as input, the analyzer will either start giving wrong output or miss out some of them.

Code Snippet:

```
/*
This also comes forward from the code snippet given above for the verb root search. Thus just the else function for the statement made for the verb root search can suffice the need for the prefix and suffix search.
*/
                else if (searchType.equals("suffix")){
                        //initializing the else function on the verb root search statement.
                        search_affix("suffix", inputWord, SUFFIXES);
                        //searching the prefixes on the prefix lexicon

                        }
```

4.5.2.3.2 G-Filter

G-Filter is the final processing module of the stem. It is here that the grammatical rules not covered in the previous modules are taken care of. The rules that are applied can be classified broadly into two categories, namely those of reordering, constraints and recursivities. Below is given a summary of the rules and processes:

4.5.2.3.2.1 Reordering

The input is processed in random order and stored in the same way. To bring them in order the rules are drawn first from what is the input and then specified as mentioned in the paradigm.

This much does not take care of the right order. For example, as has been shown in section 4.3.1 above, the same key may have more than one value. There are pronominal clitics that can come both as subject and object clitic. In this case, a

simple search results in a random choice that may be wrong. To bring surety of the results, some rules have been drawn.

Ordering of the Segmented Items

Meta Rule: Follow the ordering rule as prescribed in the verb paradigm. Take the order as given in the input string

Clitics Reordering

For the clitics having more than one value, the following rules apply:

Rule A. If there is only one clitic preceding the verb root,

Take it to be SUBJ.CL by default

Comment: This rule is not yet perfect. This case may be an object clitic also. However, for perfectness, one will need a discourse analysis or the contextual information, a higher ambition to which the current undertaking of the researcher does not aspire.

Rule B. If there are two clitics preceding the verb root

Take the first one as SUBJ.CL and the second one as OBJ.CL

Comment: This is a perfect rule and takes care of the ambiguity that might arise due to dual value of a key in the prefix lexicon.

Code Snippet

```
if ( (countSCL == 1) && (countOCL==0) )
        ts = tmp;
        //if both present and OBJ.CL is to the right of SUBJ.CL
        else if (
                ((countSCL == 1) && (countOCL==1)) &&
                (tmp.indexOf("OBJ.CL") > tmp.indexOf("SUBJ.CL"))
```

```
                )
                ts = tmp;
                else {
                st = new StringTokenizer(tmp, " ");
                int c = 0;
                //loop through all the tokens
                System.out.println("tmp="+tmp);
                        while(st.hasMoreTokens()){
                        tkn = st.nextToken();
                        // if only OBJ.CL
                        if ( (countOCL==1)&& (countSCL==0)&& (tkn.indexOf("OBJ.CL")>-1)
){
                        ts = ts + " "
+tkn.substring(0,tkn.indexOf("OBJ.CL"))+"SUBJ.CL_PREF]";
                        }
                        //if two OBJ.CL
                        else if ( (countOCL==2)&& (countSCL==0)&&
(tkn.indexOf("OBJ.CL")>-1) ){
                        if (c==1)
                        ts = ts + " "
+tkn.substring(0,tkn.indexOf("OBJ.CL"))+"SUBJ.CL_PREF]";
                                else
                                ts = ts + " " +tkn;
                        }
                        //if two SUBJ.CL
                        else if ( (countOCL==0)&& (countSCL==2)&&
(tkn.indexOf("SUBJ.CL")>-1) {
                        c++;
                        if (c==2)
                        ts = ts + " "
+tkn.substring(0,tkn.indexOf("SUBJ.CL"))+"OBJ.CL_PREF]";
                                else
                        ts = ts + " " +tkn;
                }
                        //one each and in reversed order
                        else if (
                                ((countOCL==1)&& (countSCL==1)) &&
```

```
                                    (tmp.indexOf("SUBJ.CL")>tmp.indexOf("OBJ.CL")) &&
                                    ((tkn.indexOf("SUBJ.CL")>-1)||(tkn.indexOf("OBJ.CL")>-1))
                            ){
                    c++;
        if (c==1)
        ts = ts + " " +tkn.substring(0,tkn.indexOf("OBJ.CL"))+"SUBJ.CL_PREF]";
        else if (c==2)
        ts = ts + " " +tkn.substring(0,tkn.indexOf("SUBJ.CL"))+"OBJ.CL_PREF]";
}
```

4.5.2.3.2.2 Constraints

The input verb phrase in Great Andamanese has a limited number of prefixes and suffixes. These numbers work as constraints and the system would not analyze or miss the extra affixes to the right or left of the verb root.

Code Snippet

```
if ( type.equals("prefix") && (c == 5)  )
                                    break;
                            else if ( type.equals("suffix") && (c == 3)  )
                                    break;
```

4.5.2.3.2.3 Recursivities

As there may be ambiguities regarding the verb roots or the prefixes after the first round of processing of the input text, to handle this, the options/multiple values are again sent back for better results.

As there may be more than one affix in the input word, the system must analyze all of this, one by one. For this, the system must be recursive to search for different affixes in the same lexicon. For example, for the input:

ŋuʈʰacibim

Don't come after me.

There are two occurrence of prefix and suffix each. What the system does is to search the whole of the prefixal or suffixal string and do the segmenting by matching each of them one by one with the suffix and the prefix lexicons. After one string has been found and there are more characters left, the system is prompted to search for them. The search is recursive until it finds at most five prefixes and 3 suffixes, after which the search for affixes breaks.

Code Snippet

```
//initialize variables again for the new word from the input
                    inputWord ="";
                    taggedOutput = "";
                    found = false;
                    searchType="";
                    prefix = "";
                    suffix="";
                    pre1Tagged="";
                    pre2Tagged="";
                    pre3Tagged="";
                    pre4Tagged="";
                    pre5Tagged="";
                    suf3Tagged="";
                    suf2Tagged="";
                    suf1Tagged="";
                    verbTagged="";
```

4.6 Back-End: Data files stored in UTF-8

The GAVA uses data files of three lexicons as described above. These are annotated lexicons of verb roots, prefixes and suffixes. There are more than 130 entries of verb

roots. At present there are around 60 prefixes and 20 suffixes. The list of prefixes and suffixes may be more or less as the work on the analysis of the verbs goes further. The verb roots are also more than what has been given here. But for some reasons they have not been included.

4.6.1 Webserver: Apache Tomcat 4.0

We have used Apache – Tomcat technology for the web server. Apache-Tomcat is a webserver software to develop and run Java server pages on a host. The Apache-Tomcat, earlier Tomcat, was developed by the Apache association of developers (For more visit: http://tomcat.apache.org). There are several versions of Tomcat available freely on the web. We have used the version 4.0. Tomcat 4.0 implements the servlets and Java server pages specifications from Java Software, and includes many additional features that make it a useful platform for developing and deploying web applications and web services.

4.7 How it works: An evaluation of the results

After successful testing of the verb phrases of Great Andamanese, more than 90% results were found correct. The verb types may be divided on the following basis: 1. Number of prefixes and suffixes 2. Types of Verb Roots based on the number of characters or syllables.

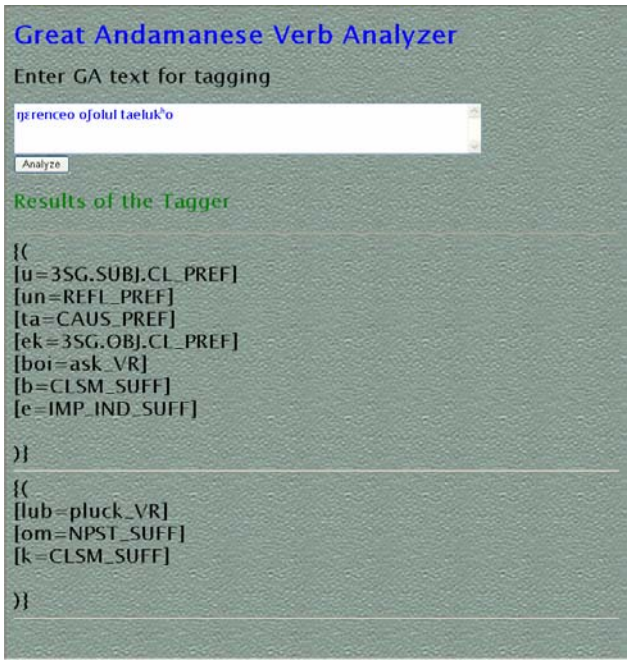
4.7.1 Differently affixed verb forms

As illustrated in the dynamics of the verb phrase in Great Andamanese, there can be a number of types of verb phrases in Great Andamanese. Some examples are given below. For more one is referred to Appendix IV where an exemplar set of verb phrases have prepared from real data examples. This appendix also tests these verb phrases on the program and gives the percentage of correct results.

4.7.1.1 Verb form with one prefix

ŋɛrenceo, oʃolul, taelukʰo

The result of the verb phrases with one syllable is almost hundred percent correct. The result is as per the expectations and that is why the result is total in favor.



Screen Shot2: Verb Root with One Prefix

4.7.1.2 Verb form with two prefixes

ŋuʈʰacibim, ŋutuncɛkʰo

Verb phrases with two prefixes are also very satisfactory. The problem arises only with the clitic decision, reflecting the doubtful condition of clitics in the grammar itself. To make it more clear, a verb phrase can have a maximum of two pronominal clitic as prefix. The problem can occur only in this case as the rule made for the ordering of the clitics has its meaning dependent more on the context than by just the position it takes in the verb phrase.

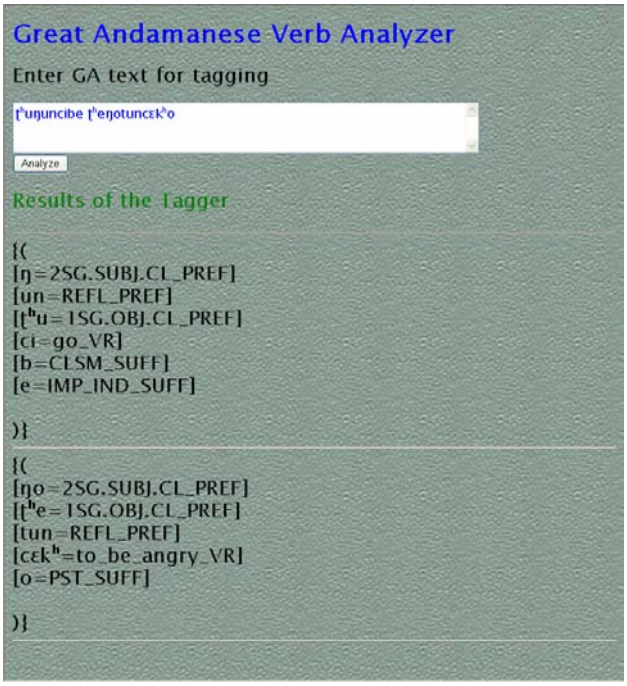
Screen Shot 3: Verb Root with Two Prefixes

However, if there are is just on clitic as prefix and any other prefix other than clitic the result will be hundred percent correct.

4.7.1.3 Verb form with three prefixes

ʈʰuŋuncibe, ʈʰeŋotuncɛkʰo

Verb phrases with three prefixes like above are also handled well by the program. The problem areas may be only those that are context dependent are cases of clitics.

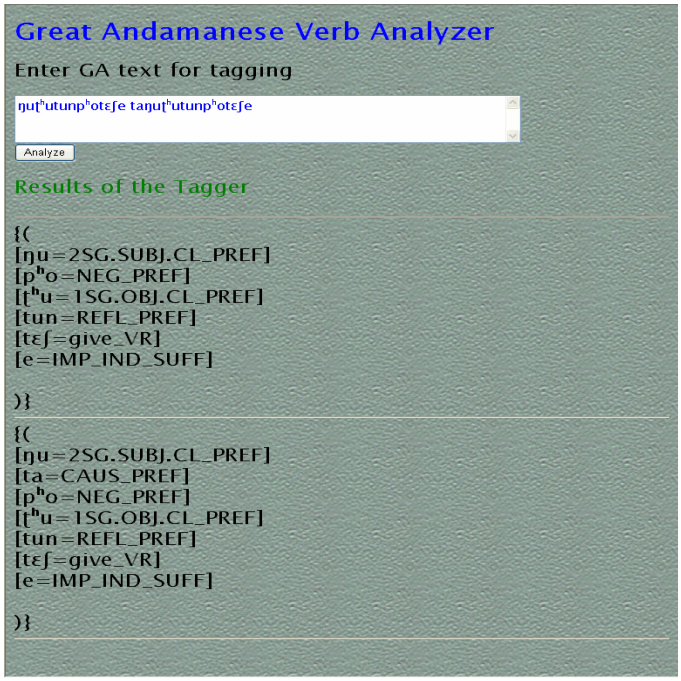


Screen Shot 4: Verb Root with Three Prefixes

4.7.1.4 Verb form with four or five prefixes

However, there are not many real cases in the data available with the researcher that are verb phrases with four or five prefixes, there can be hypothetical cases like the following where one can have a verb phrase with four prefixes.

ŋuʈʰutunpʰotɛʃe, taŋuʈʰutunpʰotɛʃe

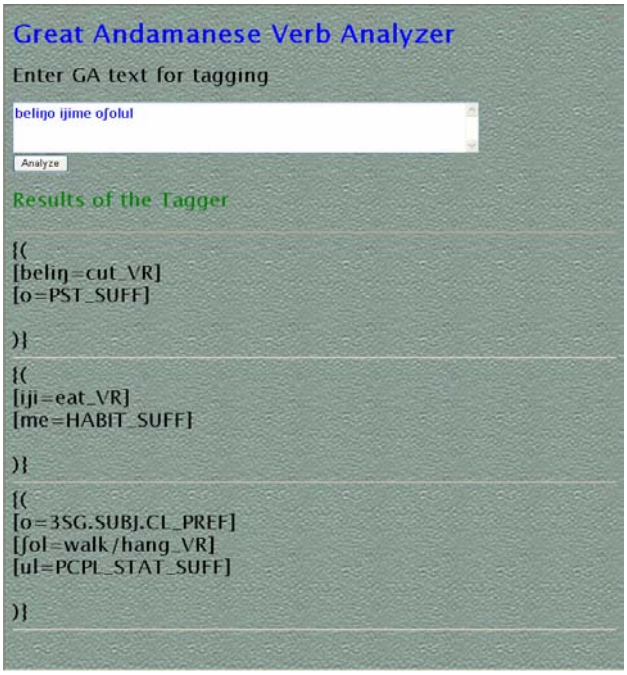


Screen Shot 5: Verb Root with Four or Five Prefixes

4.7.1.5 Verb form with one suffix

beliŋo, ijime, oʃolul

Suffixes are all handled very accurately by the program. All the prefixes are almost non ambiguous and there is only one exception to this when the prefix /-l/ comes for both the participle / stative marking and for the consonant class marking. This ambiguity has not been handled by the program as this would have demanded more of contextual study of the language and data structuring.

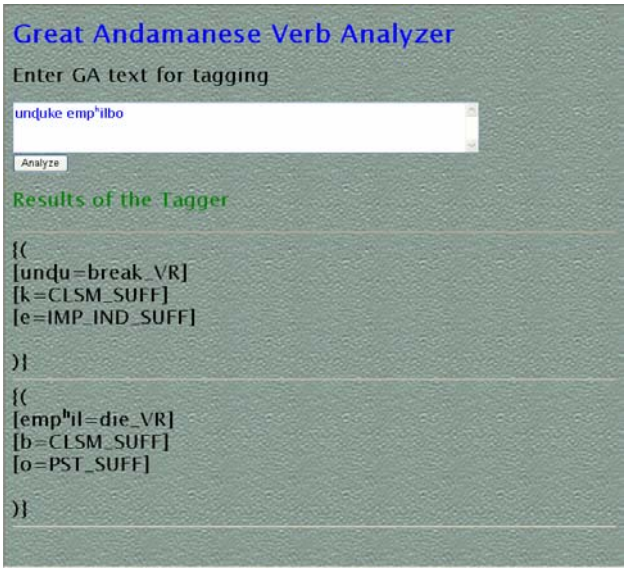


Screen Shot 6: Verb Root with One Suffix

4.7.1.6 Verb form with two suffixes

unɖuke, empʰilbo

Verb phrases with two prefixes also go very well with the analyzer. The chance of incorrectness lies only with the case of three tags given for the suffix /-l/. Other wise it goes all very fine.

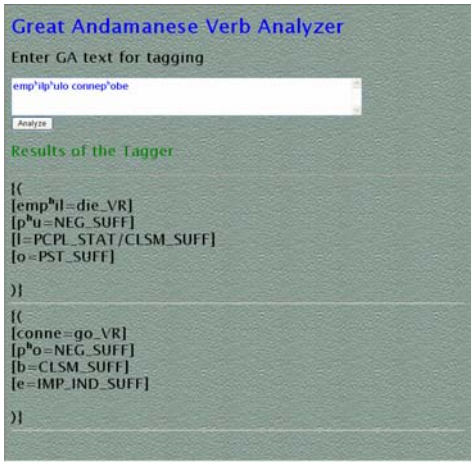


Screen Shot 7: Verb Root with Two Suffixes

4.7.1.7 Verb form with three suffixes

emp^{h}ilp^{h}ulo, connep^{h}obe

The verb phrases that can have three suffixes must contain the negative suffix /-p^{h}o/ or its allomorphs. Except for this, it will have all the structure that is already for the verb phrases with two prefixes.



Screen Shot: 8 Verb Root with Three Suffixes

4.7.2 Different Types of Verb Roots based on the Length of Characters or Syllables

As the Great Andamanese verb root does not allow more than three syllables and as the verb root lexicon suggests, the maximum number of characters available in a root is eight.

4.7.2.1 Verb Roots with three Syllables

emp^{h}orol, ereŋk^{h}ol, ravufro

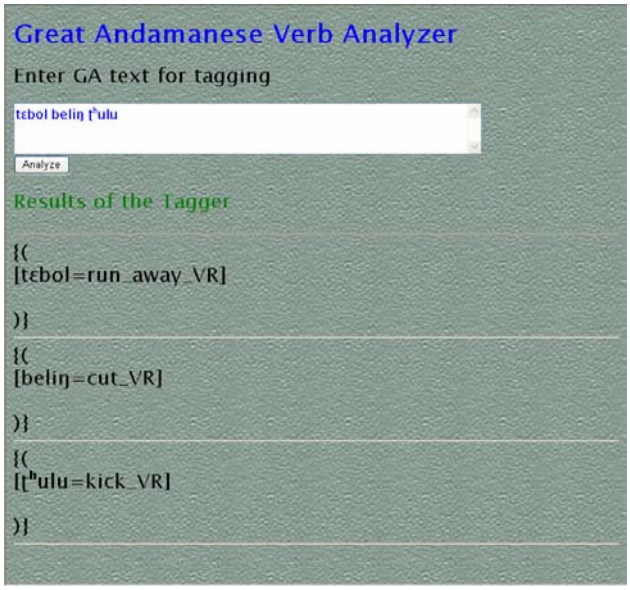
The verb roots are the core of a verb phrase. The length of a verb root also decides how it performs in the program. The larger the verb roots the longer the processing time and the more its chance of getting ambiguous. However the verb roots given in the lexicon do well with the program.

Screen Shot 9: Three Syllable Verb Root

4.7.2.2 Verb Roots with two Syllables

tɛbol, beliŋ, ʈʰulu

All the verb roots with two syllables go well with the analyzer. In the testing that has been done so far for the program, no incorrect output has been found.

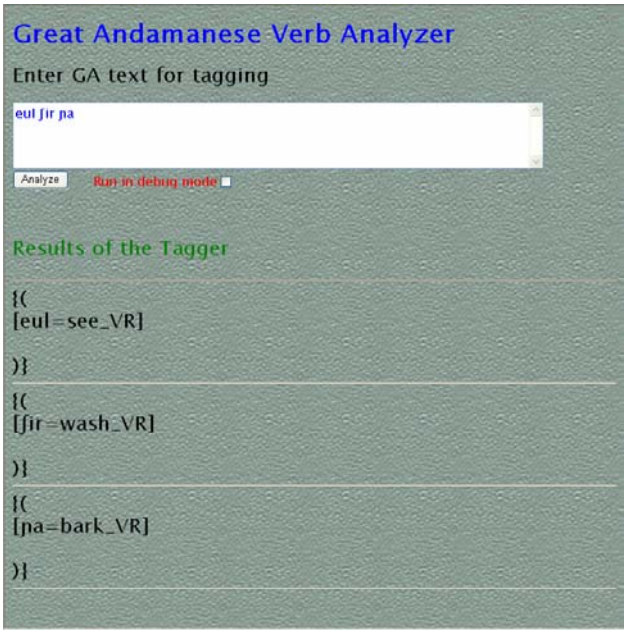


Screen Shot 9: Two Syllable Verb Root

4.7.2.3 Verb Roots with one Syllable

eul, ʃir, ɲa

Verb roots with one syllable are the best to go with the analyzer. Some verb roots of this type have character sequences that can be found in the verb roots of two or three syllables. However, this does not create any problem for the verbs with one syllable.



Screen Shot11: Mono-Syllable Verb Root

4.7.3 The Overall Result of the Analyzer

The overall result of the analyzer can be summarized only after the testing. So far, I have tested a list of verb phrases extracted from a set of model sentences. This list contains a total of 129 verb phrases. The result percentile has been extracted after the performance of the program to be up to 94%. For more on this, one should see Appendix IV where the grading scheme and other matters have also been discussed.

4.8 Handling Ambiguities/Errors

Ambiguities arise due to dual entries of the prefixes and in the verb roots that can be a sequential subset of a larger verb root such as /akaɲɔrɔ/ and /kaɲ/. This ambiguity has been handled by a rule that has been described above in the section titled *Reordering*.

- Ambiguities arising due to dual entry of the prefixes
- Ambiguities arising in the verb phrase due to the presence of a sequential subset of a larger verb root
- Ambiguities arising due to rule implementations i.e. the exceptions
- Ambiguities/Errors of Suffixes

4.8.1 Performance of the Rule

A rule for the decision of the verb root has been implemented in the program. The rule says that if there are two clitics as prefixes, the first clitic has necessarily to be a subject clitic and the second one the object clitic. This rule operates on the output given by the analyzer after the search has been completed on the verb phrases. Thus the rule superimposes the output of the rule on the output and then sends it for display.

However, this is not always fruitful. This rule also generates errors. For example a verb root preceded by a pronominal object clitic only will be interpreted as subject as it is the only clitic in the verb form.

Similarly, there are cases in which a verb form having two clitics as prefixes may not always be interpreted correctly for example /ikubeliŋ/ where the output is /[u=3SG.SUBJ.CL_PREF] [ik=3SG.OBJ.CL_PREF] [beliŋ=cut_VR]/. It is clear that the reordering has been done by the rule and this does not fit in here. To solve this problem, another rule giving priority to the input order has to be implemented so that the order of the input is not tampered with and the rule applies only in the case when the constraints are not fulfilled.

Chapter 5
Conclusion

The work carried out through the dissertation has gone several phases of analysis and the result found thus has been implemented computationally. As the title suggests, the aim of the researcher has been to develop a framework to account for the verb morphology of Great Andamanese. To fulfill this task, the researcher has undertaken the task of exploring and devising a paradigm for the verb morphology of the language under study.

The work can broadly be classified in three parts:

- Data collection/field work
- Analysis
- Computing

The method of data collection, transcription and the field situation has been described in the introductory chapter. The second chapter gives a brief introduction to the typological structure of the language. It discusses briefly the sound system, morphology and the syntax of the language. The third chapter discusses the verb morphology of the language in detail and has brought about a paradigm of the verb forms. In the fourth chapter, the computing task has been taken.

The fourth chapter describes the methods and techniques applied to bring about the framework. Computing has been the final and ultimate goal of the work done here. The ambition has been to bring about a framework that could analyze the verb morphology of the language under study. As the structure of the paradigm found for the language is unmatched by any other language that has so far come in view of the

researcher, the computing task has been daunting and required a separate approach of itself. Using the finite state automaton as the approach, implementing the paradigm to generate and analyze the verb forms of Great Andamanese gave the results that came up to be more than 90% correct.

5.1. The Work on Great Andamanese

Work on Great Andamanese has mostly been of anthropological nature. There is a lot of literature available starting from the 2nd century to the postmodern world. However, there has always been a mystery around the inhabitants of the islands as the earlier records mentioned them either with a sense of awe and fright or simply neglected them as cannibals and other derogatory terms. It was only in the modern era when the British established their penal settlement; the Andamanese also came in real and continuous contact with the inhabitants of the outer world. It was then that any kind of authentic study can be said to have started on these people.

The modern study also focused mainly on the anthropological areas of the people. The only linguistic work to have been done in the 20th century is Man's *Manual* and Manoharan's descriptive work. While the *Manual* has data on six dialects of the language family, it is more or less silent on a descriptive grammar of the language as its goal being that of a *Manual*. Manoharan's work in this sense is worth great appreciation as he describes in detail the sound system and morphology of the language. As this is an endangered language, it is also of great importance to have some data available in book form in Manoharan's descriptive study.

In recent years work on Great Andamanese was undertaken first in 2001 by Anvita Abbi who did a pilot survey of the languages of Andaman Islands. This came in time as the language has been rapidly dying. Abbi, with her papers and now a book, has brought the language and the people of the islands under present academic discussions.

The work done on Great Andamanese through this dissertation is a step forward in the existing research published on verb structure of Great Andamanese. I have tried my best to devise the verb paradigm structure. The paradigm structure takes care of the verbal suffixes namely, tense, aspect, mood and the 'class marker consonants', and the verbal prefixes, namely the pronominal clitics, reflexive, the causative/transitivizer/applicative and the occasional negative. However, the grammatical sketch done by Abbi serves as a base of this schema. This researcher has added something more to this structure and refined it to achieve the form presented in the third and implemented in the fourth chapter.

5.2. Limitations of the Dissertation

There are several limitations of the work undertaken here. While the main goal of this dissertation is to develop a framework for the verb morphology of Great Andamanese, it necessarily had to go into several other things that were either not clear or simply no work had been undertaken on this. The other things that were required to be done for the purpose can be summarized in the following way:

- Devising a solution to the dialectal problem as the term Great Andamanese stood for a conglomerate of ten languages or dialects of the Islands
- Looking for a good description of the language under study
- Deciding on the morphological type of the language
- Deciding on a discreet morphological glossing for the language
- Looking into the structure of the verb roots

The first problem was that of deciding on the term Great Andamanese. As said in the introductory chapter, in Great Andamanese, there are around ten languages or dialects. The language we studied was for sure either one of them or a mix of some of them. The name given to this has been suggested as Present Great Andamanese by Manoharan and others. However, even if someone says Great Andamanese, it would by default mean Present Great Andamanese as one can more or less find similarity in the language being spoken by the remaining people. Taking on this argument, the dissertation has used the term 'Great Andamanese' to denote the language being

spoken. The description used in the dissertation is more based on the data collected by me and similarly the glossing, morphological type other things discussed in the previous chapters take on this data.

However, the dissertation limits itself to the description of the verb morphology. It focuses more on devising a schema of the verb forms in Great Andamanese as is done in the third chapter.

5.3. Limitations of the System (the Computational Program)

This dissertation (especially the fourth chapter) basically records the processes involved in developing a computational framework for the verb morphology of Great Andamanese. The GAVA program prepared thus involves many things starting from the preparing the program to developing a user interface to developing a lexicon for the program to work on. As the ambition of this project is to develop a computational framework for the verb morphology of the language, the GAVA program does not aspire to account for all the verb roots in the language under discussion. It uses a list of about 130 verb roots. The verbal prefixes and suffixes are almost exhaustive. However it has been extracted from a limited set of data, the data set aspires to record all of the available suffixes and prefixes. Still, there may be some pronominal prefixes missing or they can be added in the lexicon to give proper results.

5.4. Potential Application

GAVA program is basically a morphological analyzer. It is highly scalable and portable system. As an NLP program, it can be used in several ways.

i) Serving as a template

GAVA program can serve as a template for further work on computing of this language or other languages having morphological systems. As the system developed is highly scalable, it can be easily adapted and extended to suit the needs of other languages as well.

ii) As a Subsystem for larger systems

GAVA can also serve as a subsystem for major NLP systems on this language or other languages with like structures. The major programs may be a general purpose parser, machine translation systems, speech recognition systems, corpus analyzers etc.

5.5. Scope of Further Research and Development

Although much research work has been done on the land and the people of the Islands, most of them relate to the anthropology or of those kinds. Serious linguistic works on the languages of these islands are scanty. That is why many linguistic areas are still to be explored. The work presented through this dissertation is focussed on the verb morphology. However, there are several other areas that it touches; a comprehensive study to them is warranted. For example, the pronominal system, the noun morphology, the syntax, the syllable structure and the discourse study are some of the interesting areas that can be explored in the language. The noun-verb dichotomy, the clitics, the deictic systems can also throw insights into the linguistic structure of the pre-Neolithic people.

Appendix I

Installation Guide

The dissertation has a CD accompanied with it. This CD contains all the necessary files and folders to install and view the results of the program prepared. The contents of the CD include the items given below. The comments following each of the item/items give a brief description of the file/folder. The folder names are given in *italics*.

A. Contents of the CD-ROM

1. Welcome.html

This is the page that opens up first when you insert the CD in the CD-Drive of your computer. The "**Start Here**" link directs you to the Index page.

2. Index.html

You have been directed to this page by the first page above. The text and six links on this page instructs and guides you how to install the application files on your computer, run them and start the web browser and test the program.

3. ExemplarVPsinGA.html

This page gives a list of verb phrases extracted from a list of basic sentences. You can dopy this phrases and test them on the Analyzer browser.

4. j2sdk-1_4_0-beta2-win.exe

This is the Java installation application kit. You can access it either from the link on the Index page or you can directly go there. Until you have it installed on your computer, you cannot run GAVA.

5. jakarta-tomcat-4.0.1.exe

This is the webserver program. Since GAVA is a Java based web-application, it needs a compatible webserver to run. Tomcat is a free webserver for Java/JSP web apps. To install this program, you can click on the link to it from the Index page or run it directly from the CD.

6. andamanese

This is the main folder of the GAVA web-application. This folder contains all the program codes and lexicons that constitute GAVA.

6.1.1. WEB-INF

This is a subfolder to the *andamanese* package above. This folder contains all the Java objects and data files. The folder structure MUST NOT be altered, else the application will not run.

6.1.1.1. classes

This folder must be a sub-folder under WEB-INF, a requirement enforced by the webserver.

6.1.1.1.1. AVTagger.class

This is the main Java object (Andamanese Verb Tagger) – a compiled version of the Java source (not included in the CD).

6.1.1.1.2. Preprocessor.class

This is the Preprocessor object - a compiled version of the Preprocessor.java file (not included in the CD).

6.1.1.1.3. verb_roots.txt

This is the lexicon of the verb roots.

6.1.1.1.4. prefix.txt

This is the lexicon of the prefixes.

6.1.1.1.5. suffix.txt

This is the lexicon of the suffixes.

6.1.2. andverbs.jsp

This is the web interface to the tagger. Receives the user request, calls the services of the tagger and displays the results.

6.1.3. 11.gif

This is a general image file. This has been used as the background of the browser page (not required by the webserver and GAVA).

7. MS Fonts

This folder contains the required font(s) to view and input the text.

7.1.1.1. ARIALUNI.ttf

7.1.1.2. Lucida Sans Unicode.ttf

These are needed only for the display. These two fonts are freely available fonts. In case if you do not have these fonts installed on your machine, you install it by downloading it from the CD. The link on the Index page will lead you to download the file.

8. Welcome_Files

These files have been generated to assist the Welcome.html.

8.1.1. filelist.xml

8.1.2. image001.gif

8.1.3. image002.gif

8.1.4. image003.gif

9. Index_Files

These files have been generated to assist the Index.html.

9.1.1. filelist.xml
9.1.2. image001.gif
9.1.3. image002.gif
9.1.4. image003.gif

10. ExemplarVPsinGA_Files

These files have been generated to assist the ExemplarVPsinGA.html.

10.1.1. filelist.xml
10.1.2. header.html
10.1.3. image001.gif

B. How to Install and View the GAVA Interface

To view the GAVA interface (the web browser), you have to insure that your computer has the following pre-installed:

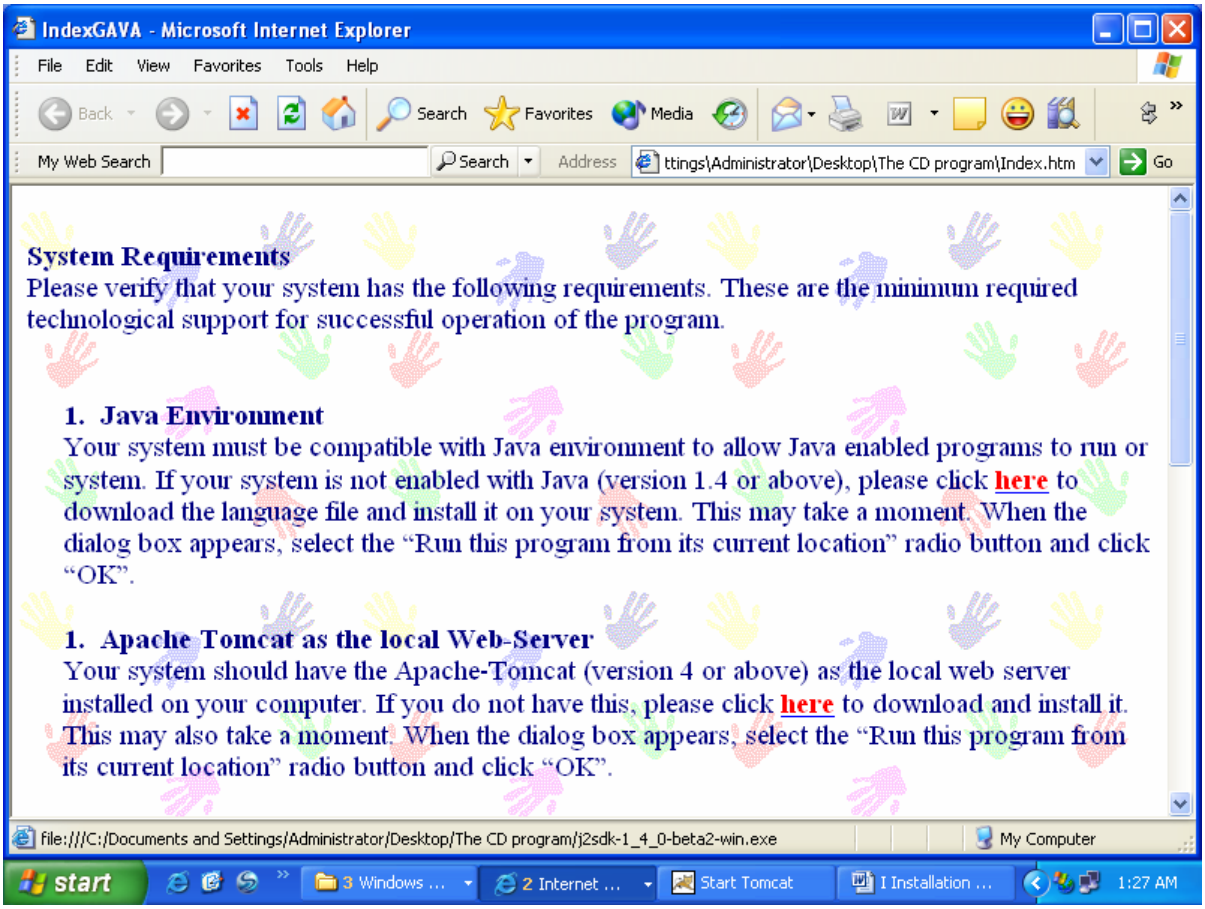
1. Java (Aache-Tomcat requires Java)
2. Apache-Tomcat as the webserver
3. Required fonts i.e. Lucida Sans Unicode or Arial Unicode MS

If you do not have any of these, all these can be installed from the CD provided with the dissertation. The Index page can be used as a guide to install and run the programs. When you insert the CD in your computer, it plays itself the following window displays on your screen:

Screen Shot 12: Start-Up Window of the CD-ROM

i. Installing the Java Development Kit and the Apache Tomcat webserver

There are six links on the Index page provided here. To install Java Development Kit click on the first **here** link as shown in the picture here:

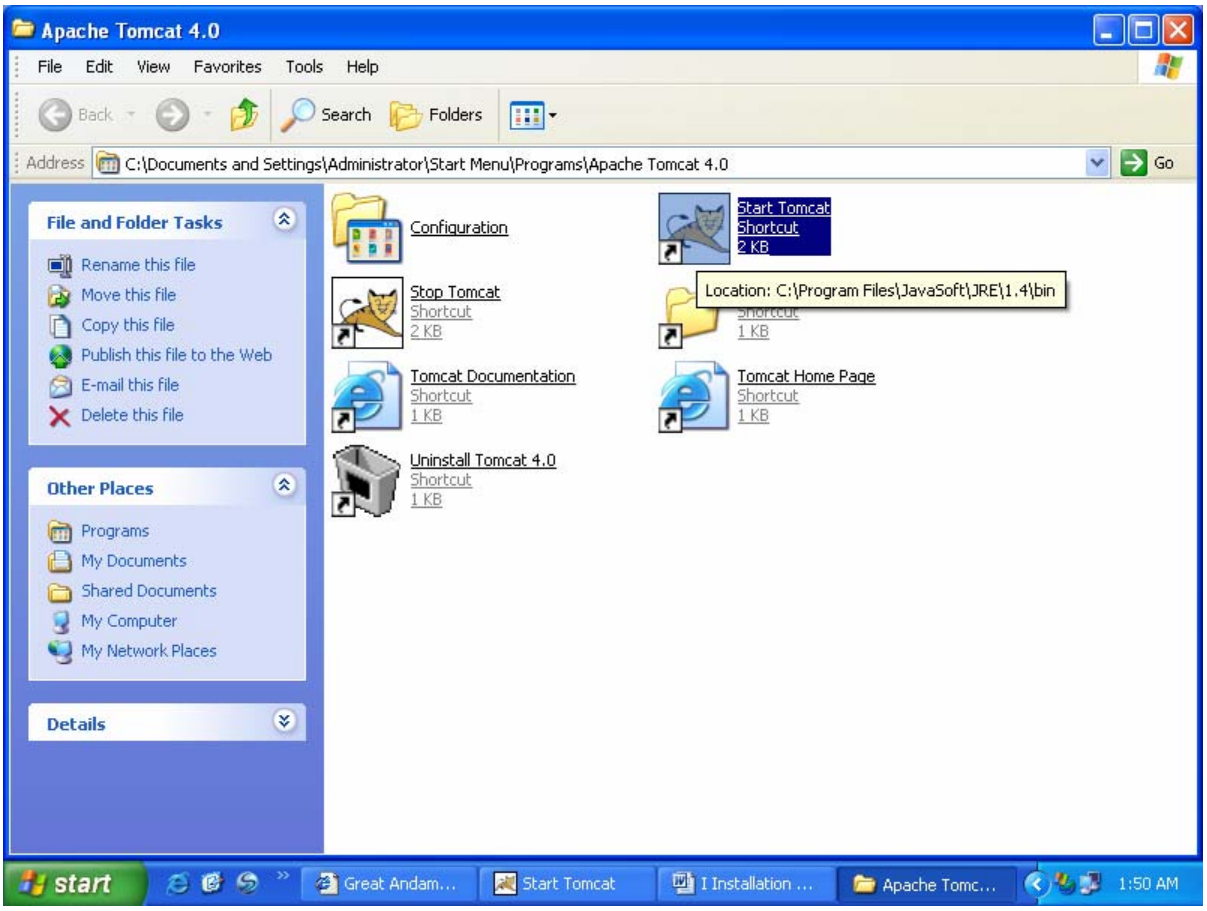


Screen Shot13: Links to the Java and Tomcat Installing Application Files

Installing these two programs is easy. Clicking on the first two links in the Index page opens a window that asks you to open/run or save the two application files. Choose to '**Open**' or '**Run**' the files. Wait for a while, there will be some other windows coming up as the *.exe* file is being installed on your computer. Keep clicking on the '**Next**' buttons and when the license agreement window opens up, choose '**Yes**'. Keep on clicking '**Next**' until the program is installed. After the Java Development Kit has been installed completely, a window will ask you to restart your computer. Choose '**Yes**' and finish the Java installation.

The Java environment has been enabled on your computer now. Now you can go to install the Apache Tomcat webserver. Here again you can use the Index page. Click on the next '**here**' button and run/open the application file. A window informing you that the local webserver is being installed using the jdk1.4 kit will open up. Click on the '**Ok**' button to continue. Keep on clicking the '**Next**' button. After the installation

is complete, a windows explorer as follows containing the content of the newly installed software will open up.



Screen Shot14: The Apache Tomcat Folder in the Start-Up Window

Make sure that you install the Java Development Kit first and then the Tomcat Setup as the latter one uses this as the compatible environment.

ii. Viewing the GAVA Interface through the web browser

After the Java development kit and the local server program have been installed, the GAVA interface can be started and viewed using a web browser. This can be done in three easy steps:

a. Step 1: Copy and Paste the Program Package

When you have these two installed on your computer, you should copy the folder named '*andamanese*' from the GAVA CD and paste it at the location <C://Program Files/Apache Tomcat 4.0/webapps>.

b. Step 2: Start Tomcat

The Apache Tomcat is the webserver on your computer. Its job is to use your machine as a host to launch the GAVA web application. The Tomcat webserver has to be started first so that the web packages can be viewed. To start the tomcat, you look for the **Start Tomcat** icon in your tomcat installation folder or by clicking start>All Programs>Apache Tomcat 4.0>Start Tomcat [1]. A command console window will open. This is the web-server application. Wait for a while for this to start up completely.

c. Step 3: Open up the GAVA Interface

After the Tomcat has started up, you can click on the URL link given on the Index page < http://andamanese:8080/localhost/andverbs.jsp>. Your browser will lead you to the GAVA interface.

Alternatively, you can type the URL <http://andamanese:8080/localhost/andverbs.jsp> on your web browser (Internet Explorer, Netscape, Mozilla, Opera etc) go to the GAVA interface.

The first page that opens up gives you a blank browser requesting you to put input text to process.

[1] If you have installed it just now, a window explorer should have opened up showing the contents of the server installed. The **Start Tomcat** and the **Stop Tomcat** buttons are shown in the contents of this window. Just click on the **Start Tomcat** button in this window. You may also like to create a Desktop shortcut to this button. To create a shortcut, right click on the Start Tomcat button, point to **Send To** and click on **Desktop (Create Shortcut)**.

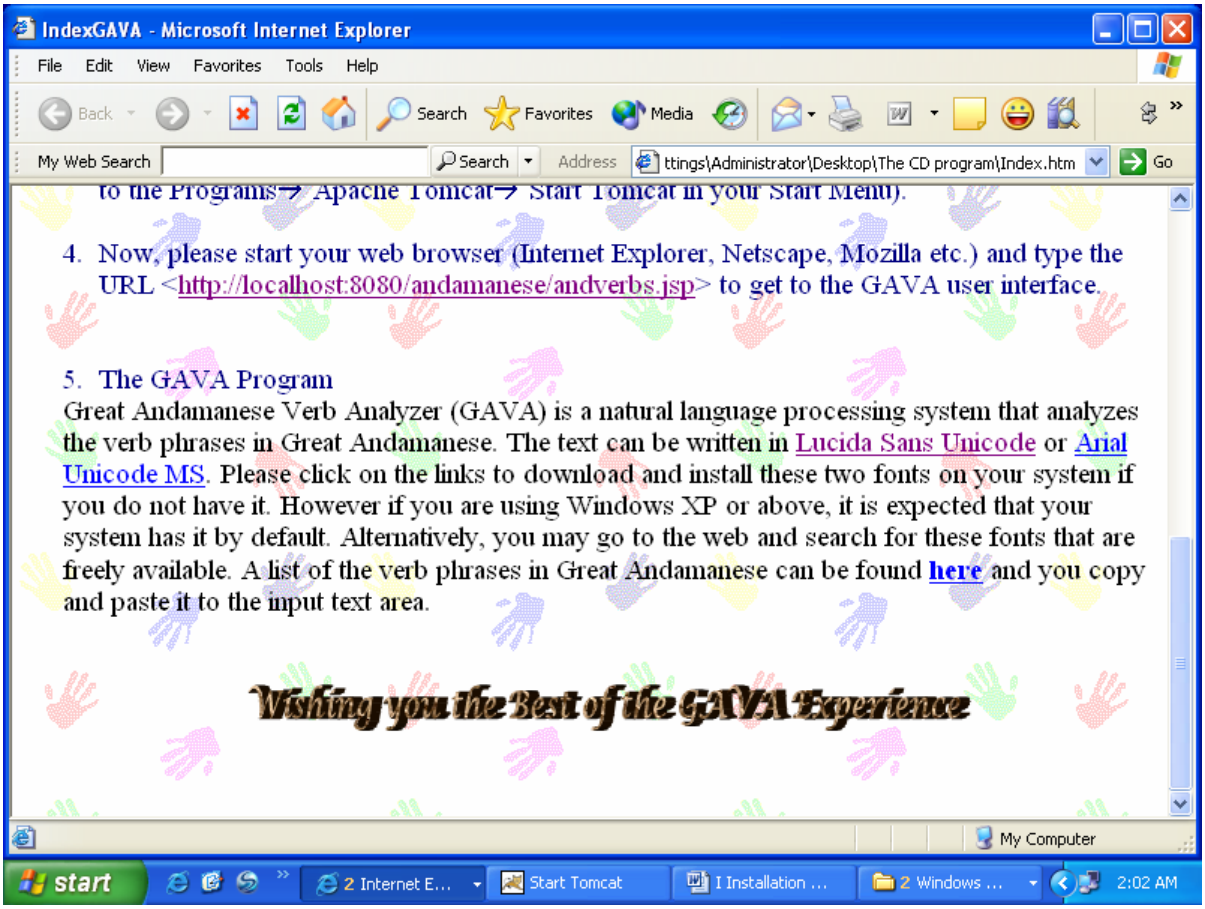
Screen Shot 15: The First Page of the GAVA Interface[2]

If you run the program in debug mode, it will take a little more time. Running the program in debug mode is useful for those who are testing the program and need to see the details of the processes undergone and the results found at every stage/module of the program.

iii. Installing the Fonts

If you do not have the necessary fonts on your computer, you may see the text given in the Exemplar Verb Phrases in Great Andamanese file as garbage! You may also not be able to input and view the text in the GAVA Interface browser. You can install them in many ways. The easiest way out will be to use the link on the Index page given here.

[2] The screen shown here with the debug button has not been used as the figures of the interface given in the fourth chapter and elsewhere.



Screen Shot16: Links to the Fonts

a. Step 1

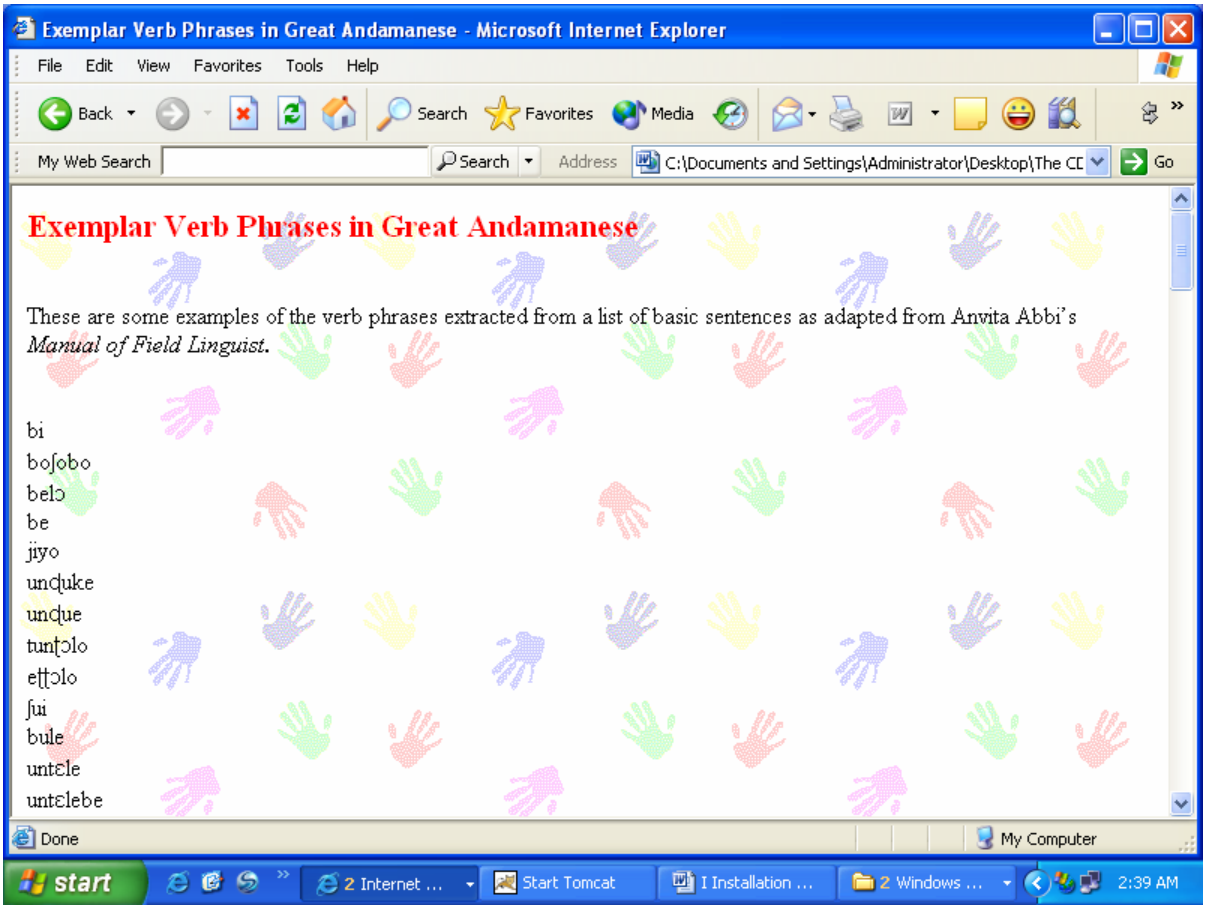
When you click on the Font names as shown in the picture above, a pop-up window asking you to open or save the file will emerge. Save the True Type File(s) on a convenient location on your computer.

b. Step 2

Now go to **Control Panel/Fonts**. Copy the fonts you have downloaded on your computer and paste it in the Fonts folder of your control panel you have just opened.

C. Running the GAVA Interface

You may not have any text to enter into the input area. A Text file has been provided here with containing a list of verb phrases. You may copy and paste them in the input text area of the GAVA interface. To access the page, click on the last link of the Index page and you will get this page



Screen Shot 17: The Exemplar Verb Phrase Window

You may copy verb phrases from here and paste it in the input text area to test the program.

Appendix II

Selective List of Sentences Used in the Dissertation

1 GA ram konabi ɟiyom
MB ram kona-bi iɟi-om
Gl Ram tendu-ACC eat-NPST
EGl Ram is eating a tendu.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-1
Rem **SIMPLE SENTENCE/ACCUSATIVE**

2 GA sita konap^{h}uŋbi ɟiyum
MB sita kona-p^{h}uŋ-bi iɟi-om
Gl Sita tendu-ripe-ACC eat-NPST
EGl Sita is eating a ripe tendu.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-1
Rem **SIMPLE SENTENCE/ADJECTIVE**

3 G.A. ət^{h}irenu ceuta konabit beliŋo
MB a-thire-nu ceu-ta kona-bit beliŋ-o
Gl SPEC-child-PL knife-INSTR tendu-ACC cut-PST
EGl The child cut the tendu fruit with a knife.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-1
Rem **SIMPLE SENTENCE/ACCUSATIVE**

4 G.A. aʃyame bəs kuttaral konabit beliŋo
MB a-ʃyam-e bəs kotr-al kona-bit beliŋ-o
Gl SPEC-Shyam- bus inside-LOC tendu-ACC cut-PST
EGl Shyam cut the tendu fruit in the bus.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-1
Rem **SIMPLE SENTENCE/ACCUSATIVE**

5 G.A. rizwane konabit beliŋel luk^{h}um beliŋo
MB rizwan-e kona-bit beliŋ-el luk^{h}um beliŋ-o
Gl Rizwan- tendu-ACC cut-PCPL finger cut-PST
EGl Rizwan cut his finger while cutting the tendu.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-1
Rem **CLAUSE COORDINATION/PARTICIPIALIZATION**

6 GA ʈ^{h}u iskule ʈhutconnep^{h}obe
MB ʈ^{h}u iskul-e ʈhut-conne-p^{h}o-b-e

Gl 1SG school-OBL 1SG- go-NEG-CLSM-IMP(NPST)
EGl I do not go to school.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-2
Rem **NEGATIVE**

7 GA kʰuro kotrakak ci-be
MB kʰuro kotra-kak ci-b-e
Gl here inside-DIREC come-CLSM-IMP
EGl Come inside.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-3
Rem **IMPERATIVE/DIRECTIONAL**

8 GA ŋu kotrakak ŋucibim
MB ŋu kotra-kak ŋu-ci-b-im
Gl 2SG inside-DIREC 2SG.SUBJ.CL-come-CLSM-PROH.NEG
EGl Do not come in.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-3
Rem **PROHIBITIVE NEGATIVE/DIRECTIONAL**

9 GA ŋaunobe
MB ŋa-uno-b-e
Gl 2SG.SUBJ.CL-sit-CLSM-IMP
EGl Please sit down
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-3
Rem **IMPERATIVE**

10 GA kʰuro kotra-kak ci-be
MB kʰuro kotra-kak ci-b-e
Gl here inside-DIREC come-CLSM-IMP
EGl Come inside.
Src Nao Jr./ 15.11.05/ MD1 Gr II T3/ Narayan C-3
Rem **IMPERATIVE/DIRECTIONAL**

11 GA ŋuʈʰi miʈʰaibi tɛʃe pʰutɛʃamo ʈʰuŋolobom
MB ŋu-ʈʰi miʈʰai-bi tɛʃ-e pʰu-tɛʃ-amo ʈʰo-ŋol-o-b-om
Gl 2SG-1SG.OBJ sweet-ACC give-IMP NEG-give-COND 1SG-cry-EPV-CLSM-NPST
EGl If you do not give me the sweets I will cry.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4
Rem **CONDITIONAL & COORDINATION**

12 GA ŋuʈʰutbatʰebamo ʈʰoŋolobom
MB ŋu-ʈʰut-batʰe-b-amo ʈʰo-ŋol-o-b-om

Gl 2SG-1SG.OBJ.CL-slap-CLSM-COND 1SG.SUBJ.CL-cry-EPV-CLSM-NPST
EGl If you slap me I will cry.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4
Rem **CONDITIONAL & COORDINATION**

13 GA oʈʰobɔyamo ʈʰui tɛrtakom
MB o-ʈʰo-bɔi-amo ʈʰu-i-tɛrta-k-om
Gl 3SG-1SG.OBJ.CL-ask-COND 1SG.SUBJ.CL-3SG.OBJ.CL- tell-CLSM-NPST
EGl If he asks me, I will tell him (the whole story).
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4
Rem **CONDITIONAL & COORDINATION**

14 GA uʈʰuncikamo uʈʰunce
MB u-ʈʰu-n-ci-k-amo u-ʈʰu-n-ce
Gl 3SG.SUBJ.CL-1SG.OBJ.CL-REFL-come-CLSM-COND 3SG.SUBJ.CL.-1SG.OBJ.CL-REFL-come-IMP
EGl If he can come, he should/must come.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-4
Rem **CONDITIONAL & COORDINATION**

15 GA aɟoe atoŋ nu təracɔre ulunciko
MB a-ɟoe a-toŋ-nu təracɔr-e eole-in-ci-k-o
Gl SPEC-Joe SPEC-Tong-PL spring-OBL see-??-go-CLSM-PST
EGl Joe and Tong went to see the spring.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-5
Rem **COORDINATION**

16 GA alepʰai ereŋkʰolebom abie pʰo
MB a-lepʰai ereŋkʰol-e-b-om a-bie-pʰo
Gl SPEC-Lephai play-EPV-CLSM-NPST SPEC-Bie-NEG
EGl Lephai will play but not Bie.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-5
Rem **COORDINATION/NEGATIVE**

17 GA kʰitani ŋiyo be
MB kʰitani ŋiyo be
Gl how 2SG.EXIST AUX
EGl How are you?
Src Nao Jr./ 08.12.05/ MD.2 BSL.3/ Narayan C-8
Rem **WH-QUESTION**

18 GA di aʃyu bi
MB di aʃyu bi
Gl 3SG who AUX
EGl Who is he?

Src Nao Jr./ 08.12.05/ MD.2 BSL.3/ Narayan C-9
Rem **YES-NO QUESTION**

19 GA at^{h}irenu taole nol
MB a-t^{h}ire-nu ta-eol-e nol
Gl SPEC-child-PL TRV-see-IND good
EGl The children looks good.
Src Nao Jr./ 08.12.05/ MD.2 BSL.3/ Narayan C-13
Rem **RELATIVIZATION/PARTICIPIALIZATION/ADJECTIVES**

20 GA a-t^{h}ire eʈɔlɔʈmobe
MB a-t^{h}ire eʈɔlɔʈmo be
Gl SPEC-child fair AUX
EGl The child is fair.
Src Nao Jr./ 08.12.05/ MD.2 BSL.3/ Narayan C-13
Rem **SIMPLE ATTRIBUTIVE SENTENCE**

21 GA amimitun at^{h}ire ta-əkaɲa
MB a-mimi-tun a-t^{h}ire ta-aka-ɲa-ɸ
Gl SPEC-mother-REFL SPEC-child TRV-3SG.OBJ.CL-eat-PST
EGl The mother fed her child.
Src Nao Jr./ 08.12.05/ MD.2 BSL.3/ Narayan C-14
Rem **CAUSATIVES/REFLEXIVE**

22 GA a-kaba a-buro girəl a-bie ta-eluk^{h}-o
MB a-kaba a-buro ik-ɟira-l a-bie ta-eluk^{h}-o
Gl SPEC-Kaba SPEC-Buro 3SGcl-tell-PCPL SPEC-Bie TRV-pick-PST
EGl Kaba got bie picked up by asking buro
Src Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-15
Rem **PERIPHRASTIC CAUSATIVES, PARTICIPLE**

23 GA aboa ameo ikɟirəl ikubeliŋ
MB a-boa a-meo ik-ɟira-l ik-u-beliŋ-ɸ
Gl SPEC-Boa SPEC-Meo 3SG.CL-tell-PCPL 3SG.SUBJ.CL -3SG.OBJ.CL-cut-PST
EGl Boa got it cut by speaking to Meo.
Src Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-15
Rem **PERIPHRASTIC CAUSATIVES, PARTICIPLE**

24 GA araːmliʃu aranʃulut^{h}u tabeno
MB araːmliʃu aranʃulut^{h}u ta-beno-ɸ
Gl sister younger brother TRV-sleep-NPST
EGl The elder sister makes her younger brother sleep.
Src Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-16
Rem **CAUSATIVES**

25 GA at^{h}ire teobi ata emp^{h}ilo
MB a-t^{h}ire teo-bi a-ta emp^{h}il-o
Gl SPEC-child crocodile-ACC 3SG.OBJ.CL-INSTR kill-PST
EGl The child got killed by the crocodile.
Src Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-17
Rem **CAUSATIVE/INSTRUMENTAL/ACCUSATIVE**

26 GA beibi tunʈɔlo
MB bei-bi tun-ʈɔl-o
Gl bottle-ACC REFL-break-PST
EGl The bottle broke.
Src Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-17
Rem **REFLEXIVE/INCHOATIVE**

27 GA b^{h}arat^{h}il nu rɛfi iɟume
MB b^{h}arat^{h}-il nu rɛfe-i iɟi-me
Gl India-LOC people rice-ACC eat-HABIT
EGl Rice is eaten in India.
Src Nao Jr./ 08.12.05/ MD.2 BSL.4/ Narayan C-18
Rem **HABITUAL**

28 GA ebuʈ^{h}u k^{h}ude oʃolop^{h}o
MB e-buʈ^{h}u k^{h}ude oʃolo-p^{h}o
Gl 3SG-weak because walk-NEG
EGl He is weak, cannot walk.
Src Nao Jr./ 08.12.05/MD.2 BSL.4/ Narayan C-19
Rem **NEGATIVE/INCAPABILITATIVE**

29 GA akaːʈa akaɲabom
MB akaːʈa aka-ɲa-b-om
Gl SPEC-girl 3SG.CL-eat-CLSM-NPST
EGl The girl is eating/eats.
Src Nao Jr./ 08.12.05/ MD.2 BSL.5/ Narayan C-20
Rem **SIMPLE NON-PAST SENTENCE**

30 GA ŋu ʈ^{h}ɛ ŋɛrenceo
MB ŋu ʈ^{h}ɛ ŋ-ɛrenceo
Gl 2SG 3SG 2SG.SUBJ.CL-fight-PST
EGl You fought with me.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-23
Rem **SIMPLE SENTENCE**

31 GA u ʈʰɛreneceo
MB u ʈʰɛ-erenceo
Gl 3SG 1SG.OBJ.CL-fight-past
EGl He fought with me.
Src Nao Jr./ 08.12.05/ MD.1 Gr.16 T.1-4/ Narayan C-23
Rem **SIMPLE SENTENCE**

32 GA ʈʰoŋɛ rɔʃe
MB ʈʰo-ŋe-rɔʃe
Gl 1SG.SUBJ.CL-2SG.OBJ.CL-love-IMP(NPST)
EGl I love you.
Src Nao Jr./ 10.01.06/ SI Narayan C-24
Rem **RECIPROCAL CHECK, SIMPLE SENTENCE**

33 GA kitaːb ʈʰica tuttaraːl ɟiyo
MB kitaːb ʈʰica tuttaraːl ɟiyo
Gl book table LOC AUX
EGl Book is on the table.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-25
Rem **LOCATIVE**

34 GA baksa tuttaraːl ʈʰutɟulu-be
MB baksa tuttaraːl ʈʰut-julu be
Gl box LOC 1SG.GEN-cloth AUX
EGl My clothes are on the box.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-25
Rem **LOCATIVE**

35 GA kitab baksa tumbol ɟiyo
MB kitaːb baksa tumbol ɟiyo
Gl book box under AUX
EGl The book is under the box.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-26
Rem **LOCATIVE**

36 GA ŋaunobe
MB ŋa-uno-be
Gl 2SG-sit-IMP
EGl Sit.
Src Nao Jr./ 10.01.06/MD.1 Gr.16 T.1-4/ Narayan C-29
Rem **SIMPLE SENTENCE/IMPERATIVE**

37 GA caːy kʰudi ŋutunceiko

MB caːy kʰudi ŋu-tun-cei-k-o
Gl what for 2SG.SUBJ.CL-REFL-angry-CLSM-NPST
EGl Why are you becoming angry?
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-29
Rem **INTERROGATIVE**

38 GA ŋu ɟobi ʃɔrokamo ŋakɛrbin caypʰo
MB ŋu ɟo-be ʃɔro-k-amo ŋa-k-ɛr-beɲ cay-pʰo-ɸ
Gl 2SG sing-IMP sing-CLSM-COND 2SG-cl-GEN-throat bad-NEG-NPST
EGl You sing. If you sing your throat will not be bad.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-30
Rem **CONDITIONAL**

39 GA ŋoiʃɔro
MB ŋo- i-ʃɔr-o
Gl 2SG.SUBJ.CL-3SG.OBJ.CL-sing-PST
EGl You sing.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-30
Rem **IMPERATIVE**

40 GA pʰɔr lobuŋ be kʰulol ɟiyo
MB pʰɔr lobuŋ be kʰulol ɟiyo
Gl bamboo long AUX there AUX-EXIST
EGl Bamboo is long. (It) is there.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-30
Rem **CLASSIFIERS CHECK, SIMPLE ATTRIBUTIVE CLAUSE COORDINATION**

41 GA pʰɔr lobuŋ be ʈʰitumulilyo
MB pʰɔr lobuŋ-be ʈʰi-tumul-il ɟiyo
Gl bamboo long-AUX place-lie-PCPL AUX-EXIST
EGl The long bamboo is lying at the place.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-30
Rem **CLASSIFIERS CHECK, SIMPLE ATTRIBUTIVE SENTENCE**

42 GA akaunol akaunol olamɔ
MB aka-uno-l aka-uno-l o-lam-ɔ
Gl 3SG.CL-sit-PCPL 3SGcl-sit-PCPL 3SG.SUBJ.CL-tire-PST
EGl He got tired while sitting.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-33
Rem **REDUPLICATION CHECK, SIMPLE SENTENCE**

43 GA akaɲal u ʈʰiɟira
MB aka-ɲa-l u-ʈʰi-ɟira

Gl 3SG.CL-eat-PCPL 3SG-1SG.OBJ-speak-PST
EGl He told me while eating.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-34
Rem **PARTICIALIZATION, COORDINATION**

44 GA cyaːk ocikom kɔil tɔ u-ʈʰuncikom
MB cya-k o-ci-k-om kɔil tɔ u-ʈʰu-inci -k-om
Gl what-DIREC 3SG.SUBJ.CL-come-CLSM-NPST later EMPH 3SG.SUBJ.CL-1SG.OBJ.CL-come-CLSM-NPST
EGl Where will he go, later he will come only to me.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-35
Rem **COORDINATION**

45 GA alepʰai kanʃɔrome eole
MB a-lepʰai kanʃɔro-me eul-e
Gl 3SG-Lephai sing-HABIT see-IMP
EGl See Lephai sing.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-37
Rem **HABITUAL MOOD**

46 GA rɛfe tunta ɲa-nol acaʃiu ɟupʰo
MB rɛfe tun-ta-ɲa-nol acaʃiu iɟi-pʰo-ɸ
Gl rice REFL-TRV-eat-good whoever eat-NEG-NPST
EGl Who does not like to eat good food?
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-40
Rem **COORDINATION**

47 GA akarkʰa o rɛfiraʃue ɲyoak enoleme
MB akarkʰa o rɛfi-raʃue ɲyo-ak e-nol-e-me
Gl 3SG.CL-say-PST 3SG rice-cook home-DIREC 3SG.CL-good-EPV-HABIT
EGl He said that cooking at home is good.
Src Nao Jr./ 10.01.06/ MD.1 Gr.16 T.1-4/ Narayan C-41
Rem **COORDINATION**

48 GA ʃup macaːn tuttaral ɟiyo
MB ʃup macaːn tuttaral ɟiyo
Gl basket table on-LOC be- EXIST
EGl The basket is on the table.
Src Nao Jr./15-11-2005/MD-1, G-12 T-4 CIILQ-1/NarayanA-61
Rem **LOCATIVE**

49 GA ceo ʃup kutraːl ɟiyo
MB ceo ʃup kutraːl ɟiyo
Gl knife basket inside-LOC be-EXIST

EGl The knife is inside the basket.
Src Nao Jr./15-11-2005/MD-1, G-12 T-4 CIILQ-1/NarayanA-61
Rem **LOCATIVE**

50 GA erŋolka biu cul ɟiyu
MB erŋolka biu cul ɟiyu
Gl matchbox candle near be-AUX
EGl The matchbox is near the candle.
Src Nao Jr./15-11-2005/MD-1, G-12 T-4 CIILQ-1/NarayanA-61
Rem **LOCATIVE**

51 GA erŋolka biu tacakʰul ɟiyu / be
MB erŋolka biu tacakʰul ɟiyu / be
Gl matchbox candle away from be-EXIST/AUX
EGl The matchbox is away from the candle.
Src Nao Jr./15-11-2005/MD-1, G-12 T-4 CIILQ-1/NarayanA-62
Rem **LOCATIVE/DEMONSTRATIVE**

52 GA o ʈɔŋ tumbol araːŋkaraːpʰːttalɛ
MB o ʈɔŋ tumbol araːŋkaraːpʰiːttalɛ
Gl 3SG tree under-LOC take rest-IMP
EGl He takes rest under the shadow of the tree.
Src Nao Jr./15-11-2005/MD-1, G-12 T-4 CIILQ-1/NarayanA-62
Rem **LOCATIVE**

53 GA ʃup macaːn tumbol ɟiyo
MB ʃup macaːn tumbol ɟiyo
Gl basket table under be-EXIST
EGl The basket is under the table.
Src Nao Jr./15-11-2005/MD-1, G-12 T-4 CIILQ-1/NarayanA-62
Rem **LOCATIVE/DEMONSTRATIVE**

54 GA pʰirʃup ʈʰica tumbol ɟiyo
MB pʰir-ʃup ʈʰica tumbol ɟiyo
Gl came-basket table under-LOC be-AUX
EGl The dustbin is under the table.
Src Nao Jr./15-11-2005/MD-1, G-12 T-4 CIILQ-1/NarayanA-62
Rem **LOCATIVE**

55 GA akacopʰonotɲyo tɛrcɔkʰel ʈʰutɲyo ɟiyo
MB akacopʰon-ka-ot-ɲyo tɛr-cɔkʰel ʈʰutɲyo ɟiyo
Gl doctor-GEN(hindi)-GEN-house infront of 1SG-GEN-house be-AUX
EGl My house is in front of the doctor's house (health centre).

Src Nao Jr./15-11-2005/MD-1, G-12 T-4 CIILQ-1/NarayanA-62
Rem **LOCATIVE**

56 GA ʈʰu ʃiro tutpʰulol ʈʰoʈoya
MB ʈʰu ʃiro tutpʰulol ʈʰo-ʈoya-ɸ
Gl 1SG sea at the coast of 1SG.SUBJ.CL-stand-NPST
EGl I am standing at the sea.
Src Nao Jr./15-11-2005/MD-1, G-12 T-4 CIILQ-1/NarayanA-62
Rem **LOCATIVE**

57 GA ɲyo tɛrcɔkʰel ʈoʈoya
MB ɲyo tɛrcɔkʰel ʈoʈoya
Gl hosue in front of stand
EGl I am standing before my house.
Src Nao Jr./15-11-2005/MD-1, G-12 T-4 CIILQ-1/NarayanA-63
Rem **LOCATIVE/DEMONSTRATIVE**

58 GA ʈʰuʈɲyo ʈɔŋ tutbɔl ɟiyo
MB ʈʰut-ɲyo ʈɔŋ tutbɔl ɟiyo
Gl 1SG.GEN-house tree behind-LOC be-EXIST
EGl I am standing before my house.
Src Nao Jr./15-11-2005/MD-1, G-12 T-4 CIILQ-1/NarayanA-63
Rem **LOCATIVE/DEMONSTRATIVE**

59 GA ŋut konabi beliŋem
M.B. ŋut kona-bi beliŋ-em
Gl 2SG tendu-ACC cut-PROH.NEG
EGl Do not cut tendu.
Src Peje/22.12.05/MD-VI-Gr.5-SI12PEJEVCLSCHK/NarayanC-75
Rem **PROHIBITIVE NEGATIVE**

60 GA ŋu ikjiram
MB ŋu ikjira-m
Gl 2SG say-PROH.NEG
EGl Do not speak.
Src Peje/22.12.05/MD-VI-Gr.5-SI12PEJEVCLSCHK/NarayanC-80
Rem **PROHIBITIVE NEGATIVE**

61 GA kʰiderbi itborɛ
MB kʰider-bi it-bor-ɛ
Gl coconut-ACC scratch-IMP
EGl Scratch the coconut.
Src Peje/22.12.05/MD-VI-Gr.5-SI12PEJEVCLSCHK/NarayanC-95
Rem **IMPERATIVE**

62 GA tekʰamo ʈʰua kʰilil ʈʰɛŋobe
MB tekʰamo ʈʰua kʰilil ʈʰe-eŋo-be
Gl now 1SG here 1SG-come-CLSM-IND
EGl Just now, I came here.
Src Nao Jr./15-11-2005/MD-1, G-12 CIILQ1 T-3/NarayanA-57
Rem **ADVERBIAL CLAUSE**

63 GA u bazaːre akabilbe
MB u bazaːr-e aka-bi-l-b-e
Gl 3SG market-LOC 3SG.SUBJ.CL-be-PCPL-CLSM-IND
EGl He is always in the market.
Src Nao Jr./15-11-2005/MD-1, G-12 CIILQ1 T-3/NarayanA-60
Rem **SIMPLE EXISTENTIAL SENTENCE**

64 G.A. unne tujulu ukutʰu
MB unne tujulu ukutʰu
Gl 3SG-PL.OBJ.CL early 3SG.SUBJ.CL-reach
EGl He reached before them.
Src Nao Jr./15-11-2005/MD-1, G-12 T-3 CIIL Q-4/NarayanA-58
Rem **ORDINAL/ADJECTIVE.**

65 G.A. lɛcbik tɛrtole
MB lɛc-bik tɛrtol-e
Gl arrow-ACC shoot-IMP
EGl Shoot arrow!
Src Peje/22.12.05/MD-6 Gr.5 SI12PEJEVCLSCHK/NarayanC-98
Rem **IMPERATIVE**

Sentences from Great Narrative of Phertajido[1]

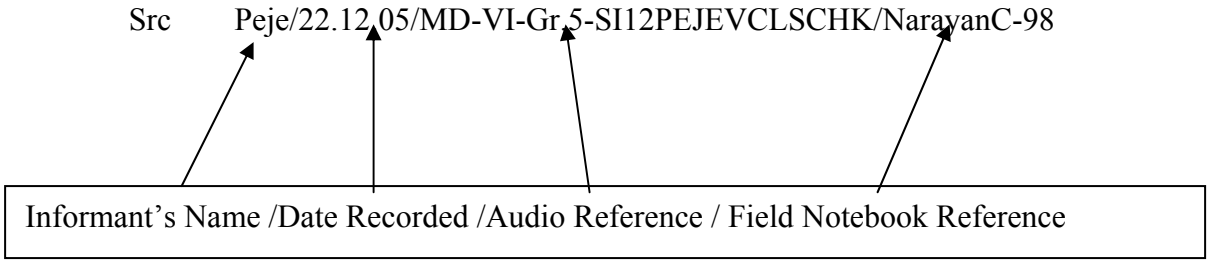
66 GA bo lɛcik kacil o lɛcit cɔŋ minotəracɛʈʰul
MB bo lɛc-ik ci-l o lɛc-it cɔŋ-ɸ mino-təra-cɛʈʰ-ul
Gl more arrrow-ACC go-PCPL 3SG arrow-ACC find-PST potato-GEN-root-LOC
EGl He went after more of the arrows (and) he found arrow in the root of a potato.
Src Nao Jr./21/01/2006 *The Great Narrative of Phertajido* Narayan A-146
Rem **COORDINATION**

67 GA kʰudi bo o lɛcik tɛrtola eka ʈʰitbolo
MB kʰudi bo lɛc-ik tɛrtol-a eka-ʈʰitbol-o

[1] This data was recorded from the memory voice recorder that is why there is no entry of the MD kind. This data has been transferred from the memory chip and is available on a separate CD.

Gl 3SG and/more arrow-ACC shoot-PST 3SG.CL-search-PST
EGl He searched for more of the shot arrows.
Src Nao Jr./21/01/2006 *The Great Narrative of Phertajido* Narayan A-147
Rem **COORDINATION**

Note
The Source reads like this:
The First entry refers to the name of the informant
Second entry is for the date recorded
The third entry refers to the audio data: **M**ini **D**isc Number, **Gr**oup Number/Name and the **T**rack Number/Name
The fourth refers to the field notebook number and the page from where the data has been taken



Appendix III

Lexicons

Lexicon A: The Verb Roots <verbroots.txt>

empʰorol=turn_VR

kaɲyɔrɔ=come_frequently_VR

kaɲɔrɔ=come_frequently_VR

ereŋkʰol=play_VR

ravufro=winnow_VR

ektɛrʈɔ=throw_VR

untɛle=call_with_happiness_VR

empʰil=die_VR

bɔkʰum=know_(neg)_VR

tabiŋo=think_VR

aratta=convince_VR

ekakʰu=open_VR

embele=overflow_VR

akaile=return_VR

tɛrtʰu=take_out_VR

raliʃo=finish_VR

bɔrɔtʰ=fall_VR

ɛrence=fight_VR

conne=go_VR

cɔnne=go_VR

rɛpʰo=climb_tree_VR

erŋol=write_VR

itpʰu=cut_VR

tɛrta=tell_VR

utlub=open_VR

mekʰu=bloom_VR

birəŋ=redden_VR
tɛbol=run_away_VR
erteɖ=see_VR
raʃui=cook_VR
beliŋ=cut_VR
elukʰ=pick_(caus)_VR
ʈʰibi=live_VR
bereŋ=pour_VR
ʃerep=cut_VR
rapʰo=cut_VR
ʈʰulu=kick_VR
kʰole=laugh_VR
ektɛr=push_VR
ipʰil=throw_VR
ɛʃilo=shake_VR
ka:ra=rise_VR
tɛrto=shoot_arrow_VR
batʰe=slap_VR
rokʰo=ready,_to_get_VR
bilup=remember_VR
boʃuʈ=hit_VR
olam=tire_VR
tʰud=pierce_VR
belɔ=aux-clsm-pst_VR
boʃo=beat_VR
eban=make_VR
biŋo=hear_VR
ɖuoc=hear_VR
eule=see_VR
meli=return_VR

bitʰ=sink_VR
jiyo=stay/ebb/AUX_EXIST/VR
koin=wake_up_VR
cɛkʰ=to_be_angry_VR
ʈɔpʰ=bathe_VR
ʃuɲe=blow,_of_nose_VR
ʈɔl=break(intr.)_VR
unɖu=break_VR
buli=defecate_VR
juvu=fly_VR
emfe=jump_VR
inci=go_VR
tɔle=mix_VR
ralɛ=moonset_VR
bele=overflow_VR
ʈɛnɔ=pull_VR
cokʰ=row_VR
koʈɛ=serve_food_VR
ʃimu=soak_VR
buli=take_away_VR
cɔpʰ=to_be_enough_VR
beno=sleep_VR
jira=speak_VR
ʈoya=stand_up_VR
kɛle=stay_VR
lele=swing_VR
emaʈ=run_VR
coŋ=get/find_VR
cɔŋ=get/find_VR
ʃɔr=sing_VR

noe=knit_VR

boi=ask_VR

bɔi=ask_VR

eŋo=come_VR

tʰu=come_out_VR

ŋol=cry_VR

ŋɔl=cry_VR

caʈ=do_VR

bɔl=peel_VR

tɔl=roam_around_VR

eul=see_VR

iye=catch_VR

tokʰ=close_VR

ʃui=cook/burn_VR

kaɲ=touch_VR

buʈʰ=fall_VR

iji=eat_VR

tɛʃ=give_VR

ʃol=walk/hang_VR

mok=leave_VR

muk=leave_VR

ɲyo=live_(home)_VR

rɔʃ=love_VR

oɖu=paste_VR

kʰi=pour_VR

kʰu=drink_VR

cɛr=rain_VR

bor=scratch_VR

lɛb=sweep_VR

cɔk=do_well_VR

ʃiʈ=hunt_VR

lub=pluck_VR

uno=sit_down_VR

ʈob=steal_VR

ŋɔʈ=swim_VR

ʃir=wash_VR

ɲa=bark_VR

ku=burn_VR

ɲa=eat_VR

cu=have_VR

ɖe=shut_up_VR

eb=take_VR

co=tie_VR

ie=give_VR

ci=go_VR

mo=give_VR

ie=pain_VR

be=AUX_VR

bi=AUX_VR

Total Number of Verb Roots: 135

Lexicon B: The Verbal Prefixes

ʈʰeŋut=1SG.SUBJ.CL.INCL_PREF

meŋut=1PL.SUBJ.CL.INCL_PREF

dunot=2PL.SUBJ.CL_PREF

tun=REFL_PREF

tum=REFL_PREF

ʈʰut=1SG.SUBJ.CL_PREF

ʈʰu=1SG.SUBJ.CL_PREF

ʈʰa=1SG.SUBJ.CL_PREF

ʈʰo=1SG.SUBJ.CL_PREF

ʈʰe=1SG.SUBJ.CL_PREF
mut=1PL.SUBJ.CL_PREF
ŋut=2SG.SUBJ.CL_PREF
dut=3SG.SUBJ.CL_PREF
uku=3SG.SUBJ.CL_PREF
aka=3SG.SUBJ.CL_PREF
ʈʰu=1SG.OBJ.CL_PREF
ʈʰa=1SG.OBJ.CL_PREF
ʈʰɛ=1SG.OBJ.CL_PREF
ʈʰi=1SG.OBJ.CL_PREF
ŋut=2SG.OBJ.CL_PREF
aka=3SG.OBJ.CL_PREF
pʰo=NEG_PREF
nu=3PL.SUBJ.CL_PREF
ʈʰ=1SG.SUBJ.CL_PREF
ʈʰ=1SG.OBJ.CL_PREF
ɛk=3SG.OBJ.CL_PREF
ɛr=3SG.OBJ.CL_PREF
ek=3SG.OBJ.CL_PREF
ne=3PL.SUBJ.CL_PREF
ik=3SG.SUBJ.CL_PREF
ik=3SG.OBJ.CL_PREF
it=3SG.OBJ.CL_PREF
ut=3SG.OBJ.CL_PREF
et=3SG.OBJ.CL_PREF
ta=CAUS_PREF
um=REFL_PREF
im=REFL_PREF
em=REFL_PREF
un=REFL_PREF

me=1PL.SUBJ.CL_PREF
ŋu=2SG.SUBJ.CL_PREF
ŋo=2SG.SUBJ.CL_PREF
ŋa=2SG.SUBJ.CL_PREF
ŋe=2SG.SUBJ.CL_PREF
ŋi=2SG.SUBJ.CL_PREF
ŋa=2SG.OBJ.CL_PREF
ŋe=2SG.OBJ.CL_PREF
ŋɛ=2SG.OBJ.CL_PREF
ŋi=2SG.OBJ.CL_PREF
ŋu=2SG.OBJ.CL_PREF
u=3SG.SUBJ.CL_PREF
u=3SG.OBJ.CL_PREF
o=3SG.SUBJ.CL_PREF
a=3SG.SUBJ.CL_PREF
e=3SG.SUBJ.CL_PREF
i=3SG.OBJ.CL_PREF
n=REFL_PREF
n=3PL.SUBJ.CL_PREF
ŋ=2SG.SUBJ.CL_PREF

Total Number of Prefixes: 59

Lexicon C: The Verbal Suffixes (suffix.txt)

amo=COND_SUFF
pʰo=NEG_SUFF
pʰu=NEG_SUFF
im=PROH_NEG_SUFF
em=PROH_NEG_SUFF
um=PROH_NEG_SUFF
om=NPST_SUFF
al=PCPL_STAT_SUFF

il=PCPL_STAT_SUFF

el=PCPL_STAT_SUFF

ul=PCPL_STAT_SUFF

me=HABIT_SUFF

l=PCPL_STAT/CLSM_SUFF

b=CLSM_SUFF

k=CLSM_SUFF

m=PROH_NEG_SUFF

e=IMP_IND_SUFF

o=PST_SUFF

ɔ=PST_SUFF

Total Number of Suffixes: 19

Appendix IV

An Exemplar Set of Verb Phrase in Great Andamanese

The following verb phrases have been extracted from the Basic Sentence List of Great Andamanese. However, all of the Sentences could not be given here for the obvious reasons of space as it will cover too much space. These samples have been tested and found correct on the system as on 30th June, 2006.

Source	Verb Phrase	Analysis Returned as Result by GAVA	Output
50,64,132, 139,140	bi	[bi=AUX_VR]	A+
12	boʃobo	[boʃo=beat_VR] [b=CLSM_SUFF] [o_PST_SUFF]	A+
35	belɔ	[belɔ=aux-clsm-pst_VR]	A+
45,79a,113 -4- 5,136,138, 42,143, 145	be	[be=AUX_VR]	A+
135137, 164-5-6	jiyo	[jiyo=stay/ebb/AUX_EXIST/VR]	A+
66	unɖuke	[unɖu=break_VR] [k=CLSM_SUFF] [e=IMP_IND_SUFF]	A+
67	unɖue	[unɖu_break_VR] [e_IMP_IND_SUFF])	A+
96	tunʈɔlo	[tun_REFL_PREF] [ʈɔl_break(intr.)_VR] [o_PST_SUFF]	A+
97	eʈʈɔlo	[e_3SG.SUBJ.CL_PREF] [ʈɔl_break(intr.)_VR] [o_PST_SUFF]	A+
126	ʃui	ʃui_cook/burn_VR	A+
133	bule	No VR entered	NA
66	untɛle	[untɛle_call_with_happiness_VR])	A+
72,73,74	untɛlebe	[untɛle=call_with_happiness_VR]	A+

		[b=CLSM_SUFF] [e=IMP_IND_SUFF]	
173	akaɲɔrɔ	[a=3SG.SUBJ.CL_PREF] [kaɲɔrɔ=come_frequently_VR]	A+
16	cibe	[ci=go_VR] [b=CLSM_SUFF] [e=IMP_IND_SUFF])	A+
17	ŋucibim	[ŋu=2SG.SUBJ.CL_PREF] [ci=go_VR] [im=PROH_NEG_SUFF] [b=CLSM_SUFF])	A+
26	uʈʰunce ~uʈʰunci	[n=3PL.SUBJ.CL_PREF] [u=3SG.SUBJ.CL_PREF] [ʈʰu=1SG.SUBJ.CL_PREF] [ci=go_VR]	A
26	uʈʰuncikamo	[n=3PL.SUBJ.CL_PREF] [u=3SG.SUBJ.CL_PREF] [ʈʰu_1SG.SUBJ.CL_PREF] [ci=give_VR] [k= CLSM_SUFF] [amo=COND_SUFF]	A
57	ʈʰuŋuncibe	[ŋ=2SG.SUBJ.CL_PREF] [un=REFL_PREF] [ʈʰu=1SG.OBJ.CL_PREF] [ci=go_VR] [b=CLSM_SUFF] [e=IMP_IND_SUFF]	A
175	neɲobe	[n=3PL.SUBJ.CL_PREF] [eɲo=come_VR] [b=CLSM_SUFF] [e=IMP_IND_SUFF]	A+
191	ŋuʈʰacebim ~ ŋuʈʰacibim	[ŋu=2SG.SUBJ.CL_PREF] [ʈʰa=1SG.OBJ.CL_PREF] [ci=go_VR] [im=PROH_NEG_SUFF] [b=CLSM_SUFF]	A
172	eɲo-ɸ	[eɲo=come_VR]	A
209	kaɲɔrɔke	[kaɲɔrɔ=come_frequently_VR]	A+

		[k=CLSM_SUFF] [e=IMP_IND_SUFF]	
210-11	kaɲɔrɔkɔ	[kaɲɔrɔ=come_frequently_VR] [k=CLSM_SUFF] [ɔ=PST_SUFF]	A+
60	raʃuekom ~ raʃuikom	[raʃui=cook_VR] [om=NPST_SUFF] [k=CLSM_SUFF]	A
111	raʃuiko	[raʃui=cook_VR] [k=CLSM_SUFF] [o=PST_SUFF]	A
22,23	ʈʰuŋolobom ~ ʈʰuŋolbom	[ʈʰu=1SG.SUBJ.CL_PREF] [ŋol=cry_VR] [om=NPST_SUFF] [b=CLSM_SUFF]	A
3	beliŋo	[beliŋ=cut_VR] [o=PST_SUFF]	A+
3	utpʰo ~ itpʰu	[itpʰu=cut_VR]	A+
85	ikubeliŋ	[u=3SG.SUBJ.CL_PREF] [ik=3SG.OBJ.CL_PREF] [beliŋ=cut_VR]	A
87	beliŋ-ϕ	[beliŋ=cut_VR]	A+
88, 89, 90	rapʰo	[rapʰo=cut_VR]	A+
91	ʃerep	[ʃerep=cut_VR]	A+
15	empʰilpʰulo	[empʰil=die_VR] [pʰu=NEG_SUFF] [l=PCPL_STAT/CLSM_SUFF] [o=PST_SUFF]	A+
71	empʰilbo	[empʰil=die_VR] [b=CLSM_SUFF] [o=PST_SUFF]	A+
193, 93, 94	empʰilo	[empʰil=die_VR] [o=PST_SUFF]	A+
51	ecaʈo	[e=3SG.SUBJ.CL_PREF] [caʈ=do_VR] [o=PST_SUFF]	A+
52	ecaʈobe ~ ecaʈbe	[e=3SG.SUBJ.CL_PREF] [caʈ=do_VR] [b=CLSM_SUFF] [e=IMP_IND_SUFF]	A+
28	kʰuo	[kʰu=drink_VR] [o=PST_SUFF]	A+
	unnikue ~ uni	bring-pst-drink-imp.pst**	NA

	kʰue		
149	kʰue	[kʰu=drink_VR] [e=IMP_IND_SUFF]	A+
1, 2	jiyom ~ jiyum ~ ijiom	[iji=eat_VR] [om=NPST_SUFF]	A+
7, 174, 107, 8	jiyo ~ ejiu ~ ijio	[iji=eat_VR] [o=PST_SUFF]	A+
28	ijul ~ ijiul	[iji=eat_VR] [ul=PCPL_STAT_SUFF]	A+
55	ŋaɲabo	[ŋa=2SG.SUBJ.CL_PREF] [ɲa=bark_VR] [b=CLSM_SUFF] [o=PST_SUFF]	A+
56	tʰaɲabo	[tʰa=1SG.SUBJ.CL_PREF] [ɲa=bark_VR] [b=CLSM_SUFF] [o=PST_SUFF]	A+
98	ijume ~ ijime	[iji=eat_VR] [me=HABIT_SUFF]	A+
203, 102, 103	jupʰo ~ jipʰo~ ijipʰo	[iji=eat_VR] [pʰo=NEG_SUFF]	A+
80	taakaɲa ~ taakaɲɔ	[a=3SG.SUBJ.CL_PREF] [ta=CAUS_PREF] [kaɲ=touch_VR] [ɔ=PST_SUFF]	F VR is “ɲa“
110	akaɲabom	[a=3SG.SUBJ.CL_PREF] [kaɲ=touch_VR] [om=NPST_SUFF] [b=CLSM_SUFF]	F VR is “ɲa“
182-2cf	akaɲal	[a=3SG.SUBJ.CL_PREF] [kaɲ=touch_VR] [al=PCPL_STAT_SUFF]	F VR is “ɲa“
69	ebuʈʰe	[e=3SG.SUBJ.CL_PREF] [buʈʰ=fall_VR] [e=IMP_IND_SUFF]	A+
70	buʈʰo~ buʈʰɔ	[buʈʰ=fall_VR] [ɔ=PST_SUFF]	A+
75	ibuʈʰɔ	[i=3SG.SUBJ.CL_PREF] [buʈʰ=fall_VR] [ɔ=PST_SUFF]	A+
76, 183	boʈʰɔ ~ buʈʰo ~	[buʈʰ=fall_VR] [ɔ=PST_SUFF]	A+

	buʈʰɔ		
127	bɔrɔtʰe	[bɔrɔtʰ=fall_VR] [e=IMP_IND_SUFF]	A+
183	oʃolul	[o=3SG.SUBJ.CL_PREF] [ʃol=walk/hang_VR] [ul=PCPL_STAT_SUFF]	A+
122	ekɛrenceo	[ek=3SG.SUBJ.CL_PREF] [ɛrence=fight_VR] [o=PST_SUFF]	A+
125	ʈʰɛrenceo	[ʈʰ=1SG.SUBJ.CL_PREF] [ɛrence=fight_VR] [o=PST_SUFF]	A+
123, 124	ŋɛrenceo	[ŋ=2SG.SUBJ.CL_PREF] [ɛrence=fight_VR] [o=PST_SUFF]	A+
53	coŋɔ	[coŋ=get/find_VR] [ɔ=PST_SUFF]	A+
176	coŋo	[coŋ=get/find_VR] [o=PST_SUFF]	A+
177	ŋutcoŋo	[ŋut=2SG.SUBJ.CL_PREF] [coŋ=get/find_VR] [o=PST_SUFF]	A+
117, 128	tɛʃe	[tɛʃ=give_VR] [e=IMP_IND_SUFF]	A+
185	ieke	[ie=give_VR] [k=CLSM_SUFF] [e=IMP_IND_SUFF]	A+
189, 195	ʈʰumoke	[ʈʰu=1SG.SUBJ.CL_PREF] [mok=leave_VR] [e=IMP_IND_SUFF]	F VR is *mo*
9,10	connepʰobe	[conne=go_VR] [pʰo=NEG_SUFF] [b=CLSM_SUFF] [e=IMP_IND_SUFF]	A+
31	eolencikɔ	see-go-clsm-pst** Verb Serailization	
33	ʈʰutconnepʰo	[ʈʰut=1SG.SUBJ.CL_PREF] [conne=go_VR] [pʰo=NEG_SUFF]	A+
33	dunotconne	[dunot=2PL.SUBJ.CL_PREF] [conne=go_VR]	A
34, 42	mutconne	[mut=1PL.SUBJ.CL_PREF] [conne=go_VR]	A
37	dutconnam ~	[dut=3SG.SUBJ.CL_PREF]	A

	dutconneom	[conne=go_VR] [om=NPST_SUFF]	
38,39	tʰɛŋutconnebe	[tʰɛŋut=1SG.SUBJ.CL.INCL_PREF] [conne=go_VR] [b=CLSM_SUFF] [e=IMP_IND_SUFF]	A+
40	meŋutconnebe	[meŋut=1PL.SUBJ.CL.INCL_PREF] [conne=go_VR] [b=CLSM_SUFF] [e=IMP_IND_SUFF]	A+
41, 42	ŋutconnepʰo	[ŋut=2SG.SUBJ.CL_PREF] [conne=go_VR] [pʰo=NEG_SUFF]	A+
46,48,49	ŋutconnebom	[ŋut=2SG.SUBJ.CL_PREF] [conne=go_VR] [om=NPST_SUFF] [b=CLSM_SUFF]	A+
47	ŋutconnebo	[ŋut=2SG.SUBJ.CL_PREF] [conne=go_VR] [b=CLSM_SUFF] [o=PST_SUFF]	A+
150	otconne	[o=3SG.SUBJ.CL_PREF] [conne=go_VR]	A-
194	inciko	[inci=go_VR] [k=CLSM_SUFF] [o=PST_SUFF]	A
156	Incikom	[inci=go_VR] [om=NPST_SUFF] [k=CLSM_SUFF]	A
69	ɛrʃoloke	[ɛr=3SG.SUBJ.CL_PREF] [ʃol=walk/hang_VR] [k=CLSM_SUFF] [e=IMP_IND_SUFF] [o=PST_SUFF]	A
100,101	boʃuʈ	[boʃuʈ=hit_VR]	A+
197-8-9	noe	[noe=knit_VR]	A
205	etmokom	[et=3SG.SUBJ.CL_PREF] [mok=leave_VR] [om=NPST_SUFF]	A+
206	mokom	[mok=leave_VR] [om=NPST_SUFF]	A+
207	mokɔ	[mok=leave_VR] [ɔ=PST_SUFF]	A+

44	ɲyokom	[ɲyo=live_(home)_VR] [om=NPST_SUFF] [k=CLSM_SUFF]	A+
129, 130,131	rɔʃe	[rɔʃ=love_VR] [e=IMP_IND_SUFF]	A+
121	ebano	[eban=make_VR] [o=PST_SUFF]	A+
84	taelukʰo	[ta=CAUS_PREF] [elukʰ=pick_(caus)_VR] [o=PST_SUFF]	A+
156	lub	[lub=pluck_VR]	A+
156	lubincikom	pick-go-clsm-npst** Verb Serialization	
180	itʰudo	[i=3SG.SUBJ.CL_PREF] [tʰud=pierce_VR] [o=PST_SUFF]	A
30, 105	ereŋkʰolebom ~ ereŋkʰolbom	[ereŋkʰol=play_VR] [om=NPST_SUFF] [b=CLSM_SUFF]	A
106	ereŋkʰolam ~ ereŋkʰolom	[ereŋkʰol=play_VR] [om=NPST_SUFF]	A
204	akamelil	[aka=3SG.SUBJ.CL_PREF] [meli=return_VR] [l=PCPL_STAT/CLSM_SUFF]	A+
21	tɛbolbe	[tɛbol=run_away_VR] [b=CLSM_SUFF] [e=IMP_IND_SUFF]	A+
184	tɛbolɔ	[tɛbol=run_away_VR] [ɔ=PST_SUFF]	A+
27	ulunciko	see-go-clsm-pst** Verb Serialization	NA
58	ʈʰoeola ~ ʈʰoeula	[ʈʰo=1SG.SUBJ.CL_PREF] [eul=see_VR] [o=PST_SUFF]	A
59	eolebo ~ eulebo	[eule=see_VR] [b=CLSM_SUFF] [o=PST_SUFF]	A+
77	taole ~ taeule	[ta=CAUS_PREF] [eule=see_VR]	A
184	erteɖel	[erteɖ=see_VR] [el=PCPL_STAT_SUFF]	A+
200	eole	[eule=see_VR]	A+

158	ŋoiʃɔro ~ ŋoʃɔro	[ŋo=2SG.SUBJ.CL_PREF] [ʃɔr=sing_VR] [o=PST_SUFF]	A
157	ʃɔrokamo ~ ʃɔrkamo	[ʃɔr=sing_VR] [amo=COND_SUFF] [k=CLSM_SUFF]	A
18,19	ŋaunobe	[ŋa=2SG.SUBJ.CL_PREF] [uno=sit_down_VR] [b=CLSM_SUFF] [e=IMP_IND_SUFF]	A+
178	akaunol	[aka=3SG.SUBJ.CL_PREF] [uno=sit_down_VR] [l=PCPL_STAT/CLSM_SUFF]	A+
201	akaunom	[aka=3SG.SUBJ.CL_PREF] [uno=sit_down_VR] [m=PROH_NEG_SUFF]	A
92	tabeno	[ta=CAUS_PREF] [beno=sleep_VR]	A
182-2cf	uʈʰijira	[u=3SG.SUBJ.CL_PREF] [ʈʰi=1SG.OBJ.CL_PREF] [jira=speak_VR]	A
41	ŋokɛle	[ŋo=2SG.SUBJ.CL_PREF] [kɛle=stay_VR]	A
161	ʈobo	[ʈob=steal_VR] [o=PST_SUFF]	A+
162, 163, 184	ʈobi ~ ʈob	[ʈob=steal_VR]	A
25	tɛrtakom	[tɛrta=tell_VR] [om=NPST_SUFF] [k=CLSM_SUFF]	A+
85	ikjiral	[ik=3SG.SUBJ.CL_PREF] [jira=speak_VR] [l=PCPL_STAT/CLSM_SUFF]	A
86	unnijiral	[n=3PL.SUBJ.CL_PREF] [i=3SG.OBJ.CL_PREF] [un=REFL_PREF] [jira=speak_VR] [l=PCPL_STAT/CLSM_SUFF] ***	NA

61,63	Tabiŋo	[tabiŋo=think_VR]	A+
62	Tabiŋobe	[tabiŋo=think_VR] [b=CLSM_SUFF] [e=IMP_IND_SUFF]	A+
67	ektɛrʈɔe	[ektɛrʈɔ=throw_VR] [e=IMP_IND_SUFF]	A+
82	ipʰilo	[ipʰil=throw_VR] [o=PST_SUFF]	A+
178, 179	olamo	[olam=tire_VR] [o=PST_SUFF]	A+
155	ŋutunceiko ~ ŋutuncɛkʰo	[ŋu=2SG.SUBJ.CL_PREF] [tun=REFL_PREF] [cɛkʰ=to_be_angry_VR] [o=PST_SUFF]	A+
192	ocikom	[o=3SG.SUBJ.CL_PREF] [ci=go_VR] [om=NPST_SUFF] [k=CLSM_SUFF]	A
170-1	ʃolom	[ʃol=walk/hang_VR] [om=NPST_SUFF]	A+
104, 154	oʃolopʰu ~ ʃolopʰu	[ʃol=walk/hang_VR] [pʰu=NEG_SUFF] [o=PST_SUFF]	A-
83	itʃiro	[it=3SG.SUBJ.CL_PREF] [ʃir=wash_VR] [o=PST_SUFF]	A+
118,119	itʃire	[it=3SG.SUBJ.CL_PREF] [ʃir=wash_VR] [e=IMP_IND_SUFF]	A+
120	erŋolom	[erŋol=write_VR] [om=NPST_SUFF]	A+

Total Number of Verb phrases checked: 129

Total Grades: 129*9=1161

Grade Points Obtained: 1084

Percentage of Correct Output: 1084*100/1161=93.37%

Grade	Number of Repetition	Grades Obtained
A+	93	93*9=837
A	30	30*8=240
A-	2	2*7=7

B+		
B		
B-		
C+		
C		
C-		
F/NA	4	4*0=0
Total Grades Obtained		1084

Single Star* represents errors due to taking a larger VR while a smaller string is meant as VR

Double Star** represents that it is a case of Verb Serialization and that it has not been taken into account.

Triple Star*** represents unexpected analysis of the affixes

Grading Scheme

Grading of the results is based on a 10 point scale.

If the verb root is not recognized, an F is given.

For every wrong output, one point is taken out.

One point is for correct ordering.

One point is for correct clitic decision.

A+ is given in the case when the output is 100% correct.

Appendix V

Informant Details[1]

A Select list of the Informants from which audio-linguistic data were ken

1. Boa Sr.

Sex: Female
Age: 80 years (approx)
Languages known: Bo, Jeru and Andamani Hindi
Education: NIL
Competence in Great Andamanese Language[2]: 5

Relation to other members of the tribe: She is the only living member of her family with her husband and children all dead.

Profile:

Boa Sr. is the eldest member of the Great Andamanese tribe. She is about 80 years old and looks rather fit for her age. She is also the only member in the tribe who does not have anybody in her family surviving. Her mother, To, was a Bo, and father, Renge, was a Jeru. She was married to Nao Sr., a Jeru, at an early age. Her parents', as well as her own marriage only testifies further a point made by her that in earlier times marriages used to take place between different language communities, i.e. different tribes. A preference for the same could easily be seen in most of the earlier matrimonial alliances. For example, out of six most senior members of the tribe, which we have recognized as having four different family lineages, four have had mixed parents.

Although she claims Jeru to be her mother tongue, which might well be the case, her language has quite an evident and strong influence of Bo. Nonetheless, she is the most proficient of the surviving Great Andamanese speakers and still retains a vast repertoire of songs and narratives. Many of her songs have such strong influence of Bo that most of

[1] The pictures and text in the informants was taken and adapted from the profile made for the project VOGA by Dr. Alok Das.
[2] The language competence level has been decided on a five point scale of proficiency in the language spoken by the people by Dr. Alok Das.

the other speakers of Great Andamanese today are unable to derive any much meaning from them. Little wonder then that Boa Sr., like most of the other Great Andamanese, insists that there is no similarity between Sare, Bo, Khora and Jeru.

She is presumably the richest surviving member of the Great Andamanese tribe in terms of linguistic-reservoir. Her love for life is quite evident when she says that she would love to stay in Port Blair. For a society which was not acquainted even with a barter system, it is interesting to observe that she understands the value of modern currency. Among the things she would often ask for are scissors, blades, and different biscuits (pickies). It is a treat to watch her when she bursts into laughter upon things she would herself say. Our predecessors would have very much been like her!

2. Boro

Sex: F
Age: 74 years ((approx))
Languages known: Khora, Jeru and Andamani Hindi
Education: NIL
Competence in Great Andamanese Language: 5

Relation to other members of the tribe: She is the step-mother of Peje. She has three sons – Golat, Loka and Sulu, and one daughter – Boa Jr., all married. Her husband is no more.

Profile:

Boro, female, 74 years, is the second oldest member of the tribe. Frail in health and introvert by nature, she rarely hesitates in showing her displeasure at anything and everything. Deep within, she is very soft and passionate and one needs to really get close to her to extract any information from her. You have to fine-tune her like an old Radio-set. Her frank and uninhibited gestures are a treat to follow. Her typically feminine negation reminds us that femininity is as old as human civilization.

Her mother, Bui, and father, Bie, were both from Khora tribe. She was married to a Jeru, Ilfe, as his second wife. She has three sons and a daughter, who are all married. She accepts Khora being her mother tongue, which she spoke, as she says, when she was

young. Presently she speaks Jeru, but obviously with strong MT (Khora) interference. Her language, therefore, is quite different from other surviving members of the tribe.

She is one of those females who generally are not interested in songs, etc., hence, remembers very few of them. She is although one of the best craftsperson in the tribe and keeps herself busy in making shell-jewellery most of the time. She has survived one attack of Tuberculosis and this has only made her more resilient and resolute.

3. Nao Jr.
Sex: Male
Age: 58 years (Official)
Languages known: Jeru, Sare, Bo, Andamani Hindi and a little bit of English too.
Education: Primary literacy
Competence in Great Andamanese Language: 5

Relation to other members of the tribe: He is the younger brother of the recently deceased King Jirake. He is married to Boa Jr. and has only one son – Bea.

Profile
He is the most docile and reasonable person in the community. He has also held several posts in the government machinery. Presently, he works as the Ward Boy in the community Medical Centre set up by the government in Strait Island. Earlier he worked in the Marine as a boat man. He has also traveled a lot. He has been to Chennai, Delhi and Kolkata and knows of the world more than anybody else in the community.

4. Peje
Sex: Male
Age: 58 years (approx)
Languages known: Jeru, Khora and Andamani Hindi
Education: NIL
Competence in Great Andamanese Language: 5

Relation to other members of the tribe: He is married to Noe with seven children – four sons (Jo, Irep, Phoro and Lephei) and three daughters (Tong, Kaba and Ilec).

Profile:

Peje is now one of the two eldest male members of the tribe after the sad demise of King Jirake in April 2005. He is about 58 years old; his father Ilfe was a Jeru, whereas his mother Muku was a Khora. He is also the stepson of Boro who had later married Ilfe. Born from mixed parents, Peje is linguistically most different from the rest of the tribe. He is the one who possesses a labialized [l] in his speech, which is not present in the speeches of anybody else in the tribe.
Peje is married to Noe with seven children, four sons and three daughters. Two of their children are married. Peje is the only person in the tribe who still spends most of his time hunting and gathering in the scant but virgin jungles of Strait Island.

5. Surmai
Sex: Female
Age: 48 years (approx)
Languages known: Jeru and Andamani Hindi
Education: NIL
Competence in Great Andamanese Language: 5

Relation to other members of the tribe: She is the widow of King Jirake with ten children: six sons (Ilphe, Meo, Nyaramo, Konmo, Baluba and Dec) and four daughters (Renge, Buro, Tango and Reya). She has one brother – Look and a sister – Noe.

Profile

Surmai is the wife of the recently deceased king of the Great Andamanese tribe, Jirake. She is 48 years old and one of those senior ladies of the tribe, who are tipped to become the head of the tribe in absence of the king and also of any legitimate male contender for the same. Both her parents were from Jeru tribe and her mother tongue is also Jeru only. She has ten children; six sons and four daughters. Besides, she also has a step daughter in Lico, who is from Jirake's earlier marriage. Only two of her children are married. One of her daughters, namely, Reya, has an unmarried male child. She does not want to disclose the name of the father and this is quietly accepted by the community. In absence of sufficient eligible bachelors in the tribe, most of Surmai's children have little option but to marry outside the tribe. This is not easy given the fact that Government approval is required before they could do so, and government policy, not surprisingly, is against this.

6. Noe F

Age: 45years (approx)
Languages known: Jeru and Andamani Hindi
Education: Literate
Competence in Great Andamanese Language: 5

Relation to other members of the tribe: She is the younger sister of Queen Surmai. She is married to Peje with seven children – four sons (Jo, Irep, Phoro and Lephei) and three daughters (Tong, Kaba and Ilec). Look is her younger brother.

Profile

Noe also works as an Anganwadi Sevika in Strait Island. She should be said to be a very dynamic lady who likes to present herself well before the world and get appreciation in return. She is like a boss in the small island and sees that things get done properly.

7. Lico

Sex: Female
Age: 43years.
Languages known: Jeru, Pujukwar, and Andamani Hindi
Education: Primary literacy
Competence in Great Andamanese Language: 5

Relation to other members of the tribe: She is the first child of King Jirake from his first wife. She is married to Golat and has five children: two daughters – Kobo and Lephe and three sons – Moroko, Buli and Berebe.

Profile:

Lico is the first child of Jirake from his first wife, Loka, who was from 'now-extinct' Pucikwar (Pujukar) tribe. It is important here to note that she was brought up by her foster maternal grand parents; grand father being a Pucikwar and grand mother a Jeru. Her biological grand parents were both Khora. Her mother Loka had died just after she gave birth to Lico. She is married to Golat with five children; three sons and two daughters. Only one of her children is married.

Lico presently works with Education department of the Andaman and Nicobar Islands and she is one of the smart women of the tribe.

Appendix VI

Pictographic Glimpses

A Group of Present Great Andamanese[1]

Feasting on the Turtle

Tribute to the Dead

The Mundan (Tonsuring) Ceremony

Interviewing a Young Mother

A Favourite Pastime

Nao Jr. in a Good Mood

The New Profession: Coconut Drying in the Sun

Nao Jr. Repairing a Transistor

Interviewing V. Pallyan, the Social Worker at Strait Island

A Group of Tourists on Boat

Ruins of the British Administrative HQ at Ross Island[1]

The Cellular Jail

The Cellular Jail[1]

The Punishment Ground inside the Cellular Jail[1]

Inside the Cellular Jail[1]


A View of the Jungles from the Sea

A Corner of an Island

The Silver Beach at Havelock

Sunrise in the Sea[1]

The Lone Indian Volcano[1]

[1] The pictures superscripted with 1 are properties of VOGA.

Appendix VII

References

Books

Abbi, Anvita. (ed.) 1997. *Languages of Tribal and Indigenous Peoples of India: the Ethnic Space*. Motilal Lal Banarasi Dass: New Delhi.

Abbi, Anvita. 2000. *A Manual of Field Linguistics and Structures of Indian Languages*. Lincom-Europa: Munich.

Abbi, A. 2006. *Endangered Languages of the Andaman Islands*. Lincom-Europa: Munich.

Aikhenvald, Alexandra Y. 2000. *Classifiers: A Typology of Noun Categorisation Devices*. Oxford University Press: New York.

Allen, James. 2004. *Natural Language Understanding.* Pearson Education Pvt. Ltd.: Singapore-New Delhi

Bakharia, Aneesha. 2001. *Java Server Page.* Prentice-Hall of India: New Delhi.

Bakker, Peter and Marten Mous (Eds). 1994. *Mixed Languages: 15 Case Studies in Language Intertwining*. IFOTT: Amsterdam.

Barbara J. Grosz et al. (ed.) 1990. *Readings in Natural Language Processing.* Morgan-Kaufmann: Los Altos, CA.

Bijalwan, C.D. 1897. *Indian Theory of Knowledge*. Heritage Publishers: New Delhi.

Chomsky, N. and M. Halle. 1968. *The Sound Pattern of English*. Harper and Row: New York.

Gerlach, Brigit and Janet Grizenhout (eds.) 2000. *Clitics in Phonology, Morphology and Syntax*. John Benjamins: Amsterdam.

Bybee, Joan & Revere Perkins & William Pagliuca. 1994. *The Evolution of Grammar – Tense, Aspect, and Modality in the Languages of the World.* The University of Chicago Press: Chicago.

Chakravorty, D.K.1990. *The Great Andamanese: Past and Present*. Seagull Books: Calcutta.

Chattopadhyay, Santanu. 2005. *Compiler Design*. Prentice-Hall of India: New Delhi.

Dahl, Östen.1985. *Tense and Aspect Systems*. Basil Blackwell Ltd: Oxford, UK.

Garnham, Allan. 1988. *Artificial Intelligence: An Introduction* Routledge and Kegan Paul: London, NY.

Givón, Talmy. 2001. *Syntax*. Volume I. John Benjamins Publishing Co: Amsterdam/Philadelphia.

Grishman, R. 1986. *Computational Linguistics: An Introduction*. Cambridge University Press: New York.

Johnsbaugh, Richard and Marcus Schaefer. 2004. *Algorithms.* Pearson Education Pvt. Ltd: Singapore-New Delhi.

Junior, Nao, Nao Sr., Look, M.R. Ranganath, S. Mahoharan and V.R. Rajasingh, 1994. *Andamani Bharati* in Andamanese-Hindi Bilingual Series-1. CIIL: Mysore

Jurafsky, Daniel, James H. Martin. 2000. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Englewood Cliffs, Prentice Hall: NJ.

Man, E.H. 1919. *A Dictionary of the South Andaman Language,* Indian Antiquary.

Manoharan, S. 1989. *A Descriptive and Comparative Study of the Andamanese language*. Anthropological Survey of India: Calcutta.

Matras, Yaron and Peter Bakker. 2003. *The Mixed Language Debate: Theoretical and Empirical Advances*. Mouton de Gruyter: Berlin.

Myers-Scotton, Carol.1998. A Way to Dusty Death: The Matrix Language Turnover Hypothesis. In Endangered Languages: Language Loss and Community Response. Ed. Lenore Grenoble and Lindsey Whaley. Cambridge University Press: Cambridge.

Portman, M.V [1898] 1992 (reprint). *Manual of the Andamanese Languages*. Manas Publications: Delhi.

Portman, M.V. 1898b. *Notes on the Languages of the South Andaman Group of Tribes,* Superintendent of Government Press: Calcutta.

Radcliffe-Brown, A.R. 1948. *The Andaman Islanders*. Free Press: Illinois.

Ringland, G.A. and Duce, D.A. 1988. *Approaches to Knowledge Representation: An Introduction,* Research Studies Press Ltd: Letchworth, Hertfordshire, England.

Ritchie, G. D., G. J. Russel Black and S. G. Pulman. 1992. *Computational Morphology*.Cambridge, Mass.:MIT Press.

Saeed, John I. 1997. *Semantics*. Blackwell Publishers: Oxford.

Song, Jae Jung. 1996. *Causatives and Causation: a universal-typological perspective*. Longman: Singapore.

Sowa, J.F. *Conceptual Structures:Information Processing in Mind and Machine.* Addision-Wesley: Reading, Massachusetts, California, London.

Spencer, Andrew. 1991. *Morphological Theory: An Introduction to Word Structure in Generative Grammar*. Basil Blackwell: Oxford.

Spencer, A. and A. Zwicky (eds.). 1998. *The Handbook of Morphology* Blackwell: Oxford.

Thomason, Sarah. 2001. Language Contact: An Introduction. Edinburgh: Edinburgh UP.

Theses

Gerdeman, Dale. 1995. *Unification Grammars for Parsing and Generation.* Ph.D. Thesis, Univ. of Illinois.

Jha, Girish N. 1994. *Morphology of Sanskrit Case Affixes: A Computational Analysis.* M.Phil. Thesis, JNU, New Delhi.

Papers

Abbi, Anvita. 2005. "Is Andamanese Typologically Divergent from Standard Average Andamanese". In the 6th Biennial Meeting of *Association for Linguistic Typology.* Padang, West Sumatra, Indonesia. 21-25 July.

Abbi, Anvita 2003. Vanishing Voices of the Languages of the Andaman Islands. Paper presented at the Max Planck Institute, Leipzig.

Abbi, Anvita *Forthcoming*. Vanishing Voices of the Great Andamanese: A Typological Sketch. In Anju Saxena's edited 'Endangered Languages' Benjamin and Co.

Annamalai, E. & Gnanansundaram. 2001. Andamanese: Biological Challenge for Language Reversal. In *Can Threatened Languages be Saved? Reversing Language Shift, Revisited: A 21st Century Perspective*. 309-22. Edited by Joshua Fishman. Multilingual Matters 116.

Basu, D.N. 1952. A Linguistic Introduction to the Andamanese. *Bulletin of the Department of Anthropology*, Vol. 1-2, July pp. 55- 70.

Burenhalt, Niclas. 1996. 'Deep Linguistic Prehistory with particular reference to Andamanese'. *Working Papers,* 45: 5-24. Lund University, Department of Linguistics, Lund, Sweden.

Endicott, Phillip, M. Thomas, P. Gilbert, Ch. Stringer, C. Lalueza-Fox, E. Willerslev, A.J. Hansen, A. Cooper. 'The Genetic Origins of the Andaman Islanders' *The American Journal of Human Genetics*. No. 72 (1), January 2003. Report no. 178.

Calder (1989, J. 1989. ´Paradigmatic Morphology' in Proceedings of the 4th European Chapter of the Association for Computational Linguistics (ACL 1989) Manchester pp.58-65.

Chomsky, N. 1956. *Three models for the description of language.* I.R.E. Transactions on Information Theory, vol. IT-2, no. 3: 113-24.

Evans, R. and G. Gazdar. 1996. ´DATR: A Language for Lexical Language Knowledge Representation' in *Computational Linguistics* 22(2) pp:167-216.

Ganguli, Pranab. 1966. Vocabulary of the Negritoes of Little Andaman with Grammatical Notes and Materials, *Bulletin of Department of Anthropology*, Vol. XV pp. 1-30.

Greenberg, J.H. 1971. *The Indo-Pacific Hypothesis* in Sebeok, T.A.(ed.) Current Trends in Linguistics, vol.8:Linguistics in Oceania. Mouton, The Hague and Paris, pp.807-871.

Hagelberg, Erika, Lalji Singh, K. Thangaraj, A.G. Reddy, V.R.Rao, S.C. Sehgal, P.A. Underhill, M. Pierson, I.G. Frame. 'Genetic Affinities of the Andaman Islanders. A Vanishing Human Population'. *Current Biology*,January 21, 2003:13, pp: 86-93

Himmelmann, Nicolaus P. (1998). 'Documentary and Descriptive Linguistics'. In *Linguistics* 36 (1998) pp: 161-95.

Jha, Girish N. 2004. *Generating Nominal Inflectional Morphology in Sanskrit*, SIMPLE 04, IIT-Kharagpur Lecture Compendium, Shyama Printing Works, Kharagpur, WB, March, 2004.

Jha, Girish N. 2005. Information Technology Applications for Sanskrit Lexicography: Case of Amarakosha, co auth. Sudhir Mishra, S Chandrashekar, in procedures of the *AsiaLex,* National University of Singapore, Singapore, 2005.

Jha, Girish, N. 2005. *Language Technology in India: A Survey* in CSI Communications, November, 2005.

Jha, Girish Nath and Sudhir Mishra, 2005. Identifying Verb Inflections in Sanskrit Morphology, In SIMPLE 05, IIT Kharagpor, pp. 79-81.

Kay, 2003, Martin. 2003. Introduction in *The Oxford Handbook of Computaional Linguistics*. Oxford: New York pp: xvii-xx.

Manoharan, S. 1986. Linguistic Peculiarities of Andamanese Family of Languages, in *Indian Linguistics.* 47 (1-4): 25-32.

Manoharan, S. 1983. "Subgrouping Andamanese Group of Languages", in International Journal of Dravidian Linguistics 12:1, 82-95

Manoharan, S. 1997. Pronominal Prefixes and Formative Affixes in Andamanese Language, in *Languages of Tribal and Indigenous Peoples of India*, Abbi, A. (ed.). MLBD: New Delhi pp: 457-473.

Pandit, T.N., 1976 "The Original Inhabitants of the Andaman Islands" in *Yojana* August15.

Radcliffe-Brown, A.R. 1914. *Notes on the Languages of the Andaman Islands*, Anthropos Vol. 9 pp. 36-52.

Siegel, Jeff. 1993. Dialect Contact and Koineisation. In *International Journal of Sociology of Language*. 99:105-121.

Sproat, R.W. 1992. Morphology and Computation. MIT Press: Cambridge, Mass.

Trost, Harald. 2003. Morphology in *The Oxford Handbook of Computaional Linguistics*. Oxford: New York pp.25-47.

Journals

VishwaBharat. A Ministry of Information Technology publication on Language Technology in India, Isuues: Sept, 2001; January, 2003; October, 2003; Jan-April-2004; October, 2004.

Government Documents

Awaradi, S.A. 1990. *Computerized Master Plan 1991-2021.* Andaman and Nicobar Administration, Port Blair.

Web References

Weber, George. 1998. *The Andamanese Language Family (I) A contribution to the Centenary of M.V. Portman's work.* 17/05/2006< http://www.andaman.org/BOOK/chapter33/text33.htm>

Weber, George. 1998. *The Andamanese Language Family (I) A contribution to the Centenary of M.V. Portman's work.* 17/05/2006
<http://www.andaman.org/BOOK/chapter8/text8.htm>

Lindfors, Anna-Lena, 2003. *Tense and aspect in Swahili*, Uppsala Universitet D-uppsats Institutionen för lingvistik HT 2003. 29/05/2006.
<http://www.lingfil.uu.se/ling/semfiler/Swa_TAM.pdf>

Ganguly, Mr. Justice S.S. et al. Andaman Administration. Report of the Andaman and Nicobar for the Backward Classes. 29/05/06
<http://www.and.nic.in/tw/obcPart%20I.pdf.>

Thangaraja, Kumarasamy, Gyaneshwer Chaubey, Toomas Kivisild, Alla G. Reddy, Vijay Kumar Singh, Avinash A. Rasalkar, Lalji Singh. "Reconstructing the Origin of Andaman Islanders "*Science* 13 May 2005:Vol. 308. no. 5724, p. 996. 14/05/2006
<http://www.sciencemag.org/cgi/reprint/308/5724/996.pdf>

John R. Pierce, John B. Carroll, et al., *Language and Machines — Computers in Translation and Linguistics*. ALPAC report, National Academy of Sciences, National Research Council, Washington, DC, 1966. 19/06/2006.
<http://www.nap.edu/books/ARC000005/html>

Burenhalt, Niclas. 1996. 'Deep Linguistic Prehistory with particular reference to Andamanese'. *Working Papers,* 45: 5-24. Lund University, Department of Linguistics, Lund, Sweden. 17/05/2006
<http://www.andaman.org/BOOK/reprints/burenhult/buhult-prehist/rep-burenhult.htm>

Bickel, Balthasar, Bernard Comrie and Martin Hapelmath. 200?. Maxplanck Institute of Evolutionary Anthropology, Leipzig. 10/11/2005
<http://www.eva.mpg.de/lingua/pdf/LGR04.09.21.pdf >