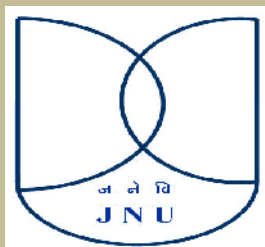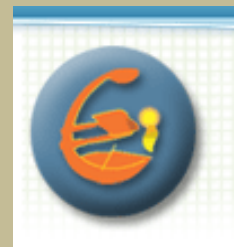Proceedings of the

Third Students' Conference of Linguistics in India
19-21 February, 2009

Centre for Linguistics,
Jawaharlal Nehru University,
New Delhi

SPONSORS

Jawaharlal Nehru University, New Delhi

CIIL, Mysore

Proceedings of
The Third Students' Conference of
Linguistics in India
(SCONLI-3)

19 – 21 February, 2009

Centre for Linguistics,
Jawaharlal Nehru University,
New Delhi

Editors:

Narayan Choudhary   **Gibu Sabu M.**

Abstract Review Committee

Faculty Members
Prof. Anvita Abbi
Prof. Vaishna Narang
Prof. Proamod K.S. Pandey
Dr. Ayesha Kidwai
Dr. Girish Nath Jha

Student Members
Narayan Choudhary
Gibu Sabu M.
Maansi Sharma
Pooja M.

Organizing Committee

Faculty Coordinator: Dr. Ayesha Kidwai
Student Coordinator: Narayan Choudhary
Student Coordinator: Gibu Sabu M.
Student Coordinator: Maansi Sharma

Members: Smita Joseph
                     Meiraba Takhelambam
                     Gopal Parihar
                     Atanu Saha
                     Ritesh Kumar
                     Sudhanshu Shekhar

Committees

Registration
- Pooja (Head) – M.Phil.
- Maansi Sharma– Ph.D.
- Rajlakshmi – M.A.
- Rajtilak– M.A.
- Madhav Gopal – M.A.

Food
- Gibu Sabu M. (Head) – Ph. D
- Meiraba Takhelambam– M.Phil.
- Atanu Saha – M.Phil.
- Bornini Lahiri – M.Phil.
- Ranjini Mazumdar– M.Phil.
- Sweta Sinha – Ph.D.

Xerox
- Ritesh Kumar (Head) – M.Phil.
- Atanu Saha – M.Phil.
- Mahesh– Ph.D.
- Prakash– M.A.
- Vivek– M.A.
- Sujoy Sarkar – M.A.

Publicity
- Narayan Choudhary (Head) – Ph.D.
- Smita Joseph – Ph.D.
- Devina Kaul– M.Phil.
- Meiraba Takhelambam– M.Phil.
- Sudhanshu Shekhar– M.Phil.

Accommodation
- Meiraba (Head - Boys) – M.Phil.
- Neha Mishra (Head - Girls) – M.Phil
- Narayan Choudhary – Ph.D.
- Maansi Sharma– Ph.D.
- Pooja M. – M.Phil.
- Radhika Gopalkrishnan – Ph.D.
- Preeti Rekha Gogoi – Ph.D.
- Atanu Saha – M.Phil.
- Mahesh M. -Ph.D.
- Ritesh Kumar – M.Phil.
- Akash Raha -M.A.
- Dev Bharati – M.A.

Hospitality
- Ranjini Mazumdar – M.Phil.
- Sabiha Hashami – M.Phil.
- Moumita Dey – M.Phil.
- Reshmi P– M.Phil.

Session Chairing
- Sabiha Hashami – M.Phil.
- Prerna Kathuria – M.Phil.
- Kulsum Mehwish – M.A.
- Deepa Thomas – M.A.
- Janini Kandhadai – M.A.
- Hima S. – M.A.
- Nimmi – M.A.
- Sujoy Sarkar – M.A.

Guest of Honour's Address

Honorable Vice Chancellor, Dean, Department Head, and Conference Organizers. I am very grateful for the invitation to say a few words at this inauguration of the Third Students' Conference of Linguistics in India. I have visited the JNU Center numerous times before, including a semester of teaching historical linguistics, during the "Monsoon Semester" 1987, the year without a monsoon, but with a great teachers' strike, and the fact that I keep returning to the Center is a clear indication of how much I respect the work done here, both by the faculty and the students. Moreover, my relationship to India is more or less that of a son-in-law (my wife is from Lucknow).

I have only had a brief time to take a look at the program for the Conference, and I am impressed by the large number of papers and the great variety of topics covered. Like the Vice Chancellor, I am especially intrigued by the paper dealing with Bhojpuri 'waalaa', and I wonder whether this is about Bhojpuri *wáalaa* or *Bhójpuri-waalaa*. I suppose this mystery will be cleared up during the Conference, and I regret that I won't be able to be there to hear the solution. I am also happy that the very first paper addresses an issue in Sanskrit grammar, with focus on the contributions of the ancient Indian grammarians. Having just taught a seminar on the teaching of historical linguistics at the CIIIL, Mysore, I note with some regret that there is no contribution in the area of historical linguistics. I hope I can inspire at least some of you to present a paper in this area at the next Students' Conference of Linguistics in India.

I am impressed by the large number of participants. This is an excellent indication that Linguistics is alive and well in India, and that a next generation is poised to take over. India, with its many languages, and its long history in some of the language families, is an ideal area for linguistic research; and its large number of endangered languages begs for intensive study of these languages before they disappear.

Again, let me thank the student organizers of the Conference, as well as Professor Ayesha Kidwai who, I understand, has played an instrumental role in helping the organizers. I am sure that the result of all of this work will be an excellent Conference.

<div style="text-align: right">

Prof. Hans Henrich Hock
University of Illinois at Urbana-Chaimpaign

</div>

Preface

The organising committee of the Third Students' Conference of Linguistics in India (SCONLI-3) bring out the proceedings of the conference with great pleasure and happiness. The Students' Conference of Linguistics in India (SCONLI) is a two day international conference organized annually by students. Initiated in the academic year of 2006-07, SCONLI strives to provide a platform to the research students to share their research with the international audience. The aim of the Conference is to promote research activities among students of Linguistics. This Conference is a step towards building a network of future generation of linguists for better, in depth and widespread work in the area of linguistics. This year (2010) it reached its fourth year and was hosted by the Department of Linguistics at Mumbai University.

This students' conference, organized fully by the students with the help of the host institution and the CIIL, Mysore, has been a great success since its start. The third conference of its series hosted at Jawaharlal Nehru University, New Delhi saw an unprecedented response from the students within India and outside. In total we had about 110 abstracts and full papers. As the SCONLI is supposed to be a two day conference followed by a day of workshops in chosen fields, we always had a limited time period. The abstract review committee, comprising of the experts in the field, along with the student members, selected 40 research works undertaken by students in various sub-areas of linguistics to be presented at the conference. With a selection ratio of just a little above to one third, the conference speaks for its quality and competitiveness itself.

We take this opportunity to thank those who were helpful in making SCONLI-3 a grand success, maintaining the momentum so that the conference goes on, attracts participation from more students, and encourages the research work undertaken by students. Dr. Ayesha Kidwai our faculty coordinator was instrumental in whatever we did. Starting from writing the proposal to both our sponsoring agencies, the CIIL, Mysore and the JNU administration, it was her wisdom and guidance in organizing such things that came to our great help. Both the sponsoring agencies were very liberal in allocating funds for this conference.

It was a challenge to provide accommodation to hundreds of students that include paper presenters and participants coming from all over India and outside. Thanks to the NCERT and the USO for providing their guest houses nearby JNU campus, we were able to accommodate all the participants.

Our next important hurdle was to host the contents of SCONLI in the website. When the conference got over, our website as hosted at our university website got dumped and we felt as if our presence and all the efforts that were put in were at a loss. Then we thought of running a website of our own. A dedicated website for SCONLI! Thus came our own website with the generic domain name of www.sconli.org. Here we hosted all the conference stuff within three days.

We would like to apologise for the long time taken for the proceedings of the conference to be published. CIIL, Mysore promised to publish the proceedings. Unfortunately that could not materialise because of some technical reasons. A lot of time has already been washed out pursuing the matter. Thanks to our generous sponsoring, we had managed to

save some bucks from the conference budget allocated for the conference. Nothing could be better than spending this money on publishing the conference proceedings. As this was a students' conference and the students being amateurs, finding an alternative publisher was also a tough job. Besides, the quality of the papers the conference received required that they see the light of their publication. Here, we would also like to regret the fact that some of the authors who presented their paper here, withdrew their papers and published them elsewhere.

About ten months after the conference was over in February, 2009, we got in touch with the current publisher willing to publish our proceedings. Shown a way out, we announced this on our website and embarked upon editing and proofreading the papers. As it was students' paper and many of the participants were writing their first research papers, amateurishness in their papers was inevitable. The proofreading and editing went through several phases, with several students from the organizing committee editing the text and the formatting. Starting from preparing the camera-ready copy of the proceedings to bringing all the papers to a particular format was a huge task and all the members, especially Ritesh Kumar and Atanu Saha, who spent time over this task need to be acknowledged at this phase.

Without the blessings of our teachers at the university and outside, the success we got in organizing this conference wouldn't have been possible. We would specially like to thank the Dr. Tanmoy Bhattacharya who also guided us on several things, starting from creating the logo for the third conference till organizing the conference. We would also thank the members who organized the previous conferences of this series, mainly Atreyee Sharma and Paroma Sanyal for their insightful comments and guidance. We thank Prof. B. B. Bhattacharya, the Vice Chancellor, Jawaharlal Nehru University and Prof. Shankar Basu, the Dean of the School of Languages, Literature and Culture Studies, JNU, who graced the inaugural session of the conference and encouraged the students. The keynote address Prof. Ramakant Agnihotri and the presidential address of Prof. Anvita Abbi, the chairperson of the department then, were enchanting speeches that encouraged the students to continue their research work and delve deeper into the research of language and linguistics with a multi-disciplinary approach.

Prof. Abbi also kept us encouraging throughout the process of going about it and made things easier for us. Last but not the least, without the help of the volunteering students of Centre for Linguistics, JNU from M.A./M.Phil./Ph.D. the success wouldn't have come our way. The enthusiasm shown by the students speaks for itself. The exhibition showcasing a panoramic view of language and linguistics in India was a great success. The posters prepared by Ritesh Kumar, Atanu Saha, Bornini Lahiri and others fetched great praises of people from all the disciplines. The pictures of these posters can be seen on the conference website www.sconli.org/sconli3 and speaks for the quality work they put into their preparation.

Enhancing students' writing ability through task oriented responses to listening exercises: the case of pre-university students in Iran

Mahan Attar
Department of English,
University of Pune, Pune-India
attarm@yahoo.com

S. S. Chopra
University of Pune,
Pune, India.
silloochopra@hotmail.com

Abstract

Writing skill, as one of the crucial aspects of communication, is an obstacle for many students. It is a complex process which requires attention to spelling, punctuation, choice of words, sentence structure and a number of other aspects. Task-oriented response to listening exercises involves the learners more communicatively in the language learning process. According to Morley (1991), there are two basic types of students' responses in listening exercises: 1) the question-oriented response model. 2) The task -oriented response model. In question-oriented response model students are asked to listen to an oral text, and then answer a series of factual comprehension questions on the content. In task -oriented response model students make use of the information provided in the spoken text, not as an end in itself but as a resource to use. This paper looks at pre-university EFL learners in Iran and describes a research project which involved the implementation of task oriented vs. question oriented responses to listening exercises, designed to enhance students' writing ability. Quantitative analysis performed on the data suggests that the task oriented responses to listening exercises is more effective than question oriented responses in promoting writing ability of Iranian students.

1 Introduction

The production and arrangement of written sentences with reasonable accuracy and coherence is a complex activity, requiring a variety of skills; and there is no general agreement among teachers regarding the methods to be used in teaching it.

The teaching of writing has undergone a tremendous change. In evolution of the teaching of writing, prior to the 1960s, teaching writing was through responding in writing to literary texts. Students read works of literature and wrote about the "themes" and the teachers trained them in techniques of literary analysis. So, correcting papers was synonymous with teaching writing.

In the 1960s, the primary concern with writing was the completed written product in a controlled composition model-product approach. Its origin seems to lie in the oral approach and the notion that language is habit formation. Writing was meant to serve as a reinforcement of the language principles and the writing task was tightly controlled in order to reduce the possibility of error. This approach did not concern the strategies and processes involved in its production. Hence, a number of forces converged to change the way composition was viewed and taught. Insights of process based inquiry began to shift teaching composition from a focus on product to a focus on process.

In the 1970s, the process approach provided a way to think about writing in terms of what the writer does (planning, revising and the like) instead of what the final product looks like (patterns of organization, spelling, grammar, etc.). Since then teachers could provide a repertoire of strategies for composing texts and encourage students to discover what they want to say during the process of writing. But, they needed to adapt a well-grounded philosophical stance to underpin their own approach to the teaching of writing. They had to choose methodologies and materials which that arose from principled decisions. Without a stand on how to promote student learning, teachers

would be forced to make ad hoc choices which may or may not be the best possible ones for the student.

2 Task-Based Teaching

In English language teaching (ELT), there exists an opinion that successful learning is influenced by appropriate methods of teaching, and task based teaching can be one of them. The emphasis on the task-based learning and teaching is reflected in much current research that studies the characteristics of different kinds of activities and tasks. It is possible that the late 1990s will be known in applied linguistics as "the age of the task" (Johnson, 2001:194). The idea of "task" is not as simple as it might seem. Many definitions and perspectives exist. According to Prabhue 1987, "*tasks are activities that require learners to arrive at an outcome from given information through some process of thought, and which allows teachers to control and regulate that process.*" (P.17).In other words students learn to use a language through applying appropriate methods and techniques of teaching. In order to make the students more competent in the communicative aspect of the language, course designers have devoted most of their time in generating exercises on teaching listening and writing in tandem.

Morley (1991) pointed out that there were two basic types of students' responses to listening exercises: 1) the question – oriented response model. 2) The task oriented response mode. In the question-oriented response model, students are asked to listen to an oral text, and then answer a series of factual comprehension questions on the content. In task-oriented response model, students make use of the information provided in the spoken text, not as an end in itself but as a resource to use. The researchers' belief was that if the effectiveness of the task-oriented response model could be proved, this kind of activity would serve as a useful supplementary material to course books in Iran, in order to create a situation in the language classroom that would improve students' writing ability.

3 Objective of the study

Since writing skill is an obstacle for many Iranian students of English as foreign language, this study is aimed at finding an effective way for enhancing students writing ability through task oriented responses to listening exercises. To achieve the purpose of this study, the following research question was proposed: "Do the Iranian pre-university students achieve more ability in writing through the task-oriented responses or through the question-oriented responses to listening exercises?"

Also, in order to find the proper response to the question, the following null hypothesis was formulated: "There is no difference between task-oriented and question oriented responses to listening exercises in enhancing the writing ability of Iranian Pre-University students.

4 Design

This study follows an experimental design that contains three characteristics: 1) experimental and control group 2) random selection of the groups and 3) a pretest-post test administration.

Participants

The subjects of the present research included 50 Iranian female pre-university students. They were selected from among 80 pre-university students on the basis of the pre-test and randomly divided into two groups: 25 students in the experimental group and 25 students in the control group.

Procedure

In order to estimate the general knowledge of the subjects, a validated Nelson test was administered to all the subjects and; then a writing pre-test (writing about pictures) that was developed by the researchers was managed. The researchers provided 10 pictures. Each subject randomly selected two pictures out of ten and wrote about them. In order to prove the validity of writing test, correlation between the final scores of writing pre-test and Nelson test scores was estimated.

Both groups participated for in ten sessions of instruction of twenty minutes each over a period of four weeks. The control group responded to listening exercises through multiple choice, fill in the blanks and true/false questions while the experimental group class responded to listening exercises based on tasks. (For example questions to write for verification, clarification and elaboration.)

Then both groups took part in writing and Nelson post-test. The aim of the writing test was to estimate the writing skill; categories of grammar, vocabulary, spelling and punctuation, and the aim of Nelson test was to estimate the validity of the writing test. In order to prove the validity of the writing test, correlation between the final scores of writing post-test and Nelson post-test scores was estimated.

Instructional Materials

The input-based component of the sessions consisted of 10 listening passages selected from the pre-university English course book. They were about issues such as, how penguins live, how we can keep a glass from cracking, etc.

There were some cards and pictures too, designed by the researchers to help the subjects both in experimental and control groups, to understand the passages better.

5 Data Analysis

The result of this experiment is provided in the following tables. The first analysis deals with the comparison of the two means of the pre-test of the control and experimental group for writing and Nelson tests. As it is shown in the following table, the t-critical is higher than the t-value and so we can conclude that the difference between the two groups is not significant and they are almost at the same level of writing ability. The result is also the same for the Nelson test.

The data derived from Nelson Pre test

	N	M	SD	T
Experimental group	25	52.5	4.73	
				0.866
Control group	25	53.7	5.05	
D.F. = 48	P = 0.05	t-critical = 2.00		

Table 1

The data derived from writing Pre test

	N	M	SD	T
Experimental group	25	44.88	15.08	
				0.264
Control group	25	43.84	12.55	
D.F. = 48	P = 0.05	t-critical = 2.00		

Table 2

The second analysis contains the data obtained from the writing and Nelson post-test scores of the two groups. As it is illustrated in the following tables, the t-value for both speaking and Nelson is higher than the t-critical and we are quite safe in rejecting the null hypothesis.

The data derived from Nelson Post test

	N	M	SD	T
Experimental group	25	65.3	10.44	
				3.001
Control group	25	57.8	6.85	
D.F. = 48	P = 0.05		t-critical = 2.00	

Table 3

The data derived from writing Post test

	N	M	SD	T
Experimental group	25	67.88	9.22	
				3.998
Control group	25	54.66	13.71	
D.F. = 48	P = 0.05		t-critical = 2.00	

Table 4

The aim of the third analysis is to compare the two means of control group on pre-test and post-test for Nelson and writing separately. The obtained t-value from match t-test is higher than the t-critical and this means that there is a significant difference between the scores of the control group for Nelson and writing on pre-test and post-test.

The data derived from Nelson
Pre and post test for control group

	N	M	SD	T
Pre test	25	53.7	5.05	
				2.405
Post test	25	57.8	6.85	
D.F. = 24	P = 0.05		t-critical = 2.064	

Table 5

The data derived from writing
Pre and post test for control group

	N	M	SD	T
Pre test	25	43.84	12.55	
				2.908
Post test	25	54.66	13.77	
D.F. = 24	P = 0.05		t-critical = 2.064	

Table 6

In analysis 4, the data obtained on writing and Nelson pre-test and post-test of experimental group is discussed. The data obtained in this comparison through match t-test shows that the t-value is higher than the t-critical and this means that there is a significant difference between the scores of the experimental group for Nelson and writing on pre-test and post-test but this difference is much higher than the control group, as shown below:

The data derived from Nelson
Pre and post test for Experimental group

	N	M	SD	T
Pre test	25	52.5	4.73	
				5.582
Post test	25	65.3	10.44	
D.F. = 24	P = 0.05		t-critical = 2.064	

Table 7

The data derived from writing
Pre and post test for Experimental group

	N	M	SD	T
Pre test	25	44.88	15.8	
				6.504
Post test	25	67.88	9.22	
D.F. = 24	P = 0.05		t-critical = 2.064	

Table 8

So, the null hypothesis was rejected and it was proved that the task-oriented responses to listening exercises are more effective than question-oriented responses in enhancing the writing ability of Iranian pre-university students.

6 Conclusion

4

Since English is a foreign language for Iranian students and also students' responses to listening exercises are question oriented, it is necessary to provide them with opportunities to use English more communicatively. So, in this study, the task oriented listening exercises were applied and compared in an experimental group with the question oriented responses in a control group. The rejection of null hypothesis would prove the effectiveness of task oriented responses in improving the writing ability. The answer to the research question is that, task oriented listening exercises increases the writing ability and the students are, not only able to respond to controlled listening comprehension questions, but also able to increase their writing ability.

So the result of this research, both emphasizes a useful way of teaching listening comprehension and an effective method to enhance the writing ability of students.

References

Bahns, J., 1995. There's More to Listening than Meets the Ear (respective review article.) System 23(3):531-547.

Canale M, Swain M. 'Theoretical bases of communicative approaches to second language teaching and testing'. Applied Linguistics (1980) 1(1):1–47.

Haitrston, M. 1982. The winds of change: Thomas Kuhn and the revolution in the teaching of writing. College Composition and Communication, 33(1), 76-88.

Hymes D. 1972. 'On communicative competence' in J. B. In: Sociolinguistics—Pride, Holmes J, eds. Harmondsworth: Penguin

Johnson, K. 2001. An Introduction to Foreign Language Teaching and Learning [M]. London : Longman.

Lynch,T .1996.Communication in the language classroom. Oxford : Oxford University Press.

Morley,J.,1991. Listening Comprehension in Second / Foreign language Instruction. Celce Murcia's book of "Teaching English as a second or Foreign Language." 2nd ed. California: Heinle and Heinle Co.

Nunan, D. 1989. Designing tasks for the communicative classroom. Cambridge : Cambridge University Press.

Nunan, D. 2004. Task-based language teaching. Cambridge: Cambridge University press Prabhu, N. S. (1987). Second language pedagogy. Oxford: Oxford University Press.

Vandergrift,L.1997. The Cinderella of Communication Strategies: Reception Strategies in Interactive Listening. Modern Language Journal. 81(4):494-505.

Wilkins, D. 1976. Notional syllabuses. Oxford: Oxford University.

Willis, J. 1996. A framework for task-based learning. Harlwo, Essex U.K: Longman/Addison-Wesley.

Comparative study of Nagpuri Spoken by Chik-Baraik & Oraon's of Jharkhand

Sunil Baraik
Senior Research Fellow (UGC)
Deptt. Of Tribal & Regional Languages
Ranchi University. Ranchi. (Jharkhand)
s_baraik_in@yahoo.com

Abstract

Nagpuri is the Mother-tongue of ChikBaraik as well as nine other tribes of Jharkhand. In this state many tribes like the Munda, the Oraon, the Kharia, the Chik Baraik and a total of thirty two tribes dwell together peacefully, all from different language groups (Austro Asiatic, Dravidian and Indo Aryan). Almost all the tribal people speak Sadani (another name for Nagpuri) with some local variation. It is used by a large section of the tribals as well as the non-tribal population either as a Mother tongue or as a lingua franca. There are different views regarding the origin and status of Nagpuri. According to the Encyclopedia Mundarica, Sadri is the language of Sadans. It is also used as a Mother-tongue among some of the Munda, Oraon and Kharia families residing in some parts of Simdega, Gumla, Lohardaga, Hazaribagh, Ranchi and Khunti Districts. Apart from Jharkhand, Nagpuri is also spoken in Assam, West Bengal, Orissa, Chhatisgarh and Bhutan where the people of Jharkhand migrated to earn their living. In those places this language is also known as Sadri, Sadani, Gawari, Nagpuri, Nagpuria and even Jharkhandy language. It is observed that almost all the Oraons know this language. It is said that Oraons were banned from using their Mother-tongue (ie; kurukh) in Chotanagpur by the kings. Hence they were forced to adopt Nagpuri as their mode of communication because it was the (Rajbhasha) official language of Chotanagpur at that time. Though the Oraons adopted this language as the means of communication, they also mixed some other tribal as well as Hindi words. Nagpuri spoken by the Chik Baraiks is very much distinct from the others. Nagpuri spoken by Chik Baraik tribe and Oraon tribe can very easily be differentiated. My paper will be based on the comparative study of Nagpuri spoken by Chik Baraiks and Oraons of Jharkhand.

7 Introduction of Jharkhand

The newly born state of Jharkhand forms the eastern part of the central tribal belt of India. It has been the homeland of different indigenous tribal, artisan and peasant communities who migrated into it in different periods of history. Through time they have evolved a common way of life where their diversities converged into a unity to a large extent.

8 The Sadans

Apart from the Munda, Oraon, Khadia, Ho, and Santhals other tribals like Ghansi, Lohra Jhora, Binjhia, Teli, Kewta, Vogta, Kumhar, Ahir ,Chik Baraik, etc is residing in Jharkhand. They all come under the Sadan nomenclature. The castes which are mentioned above are called 'Sadan' but since when these castes have been living with Mundas and Oraons is difficult to find out.

9 Chik Baraik

Chik Baraik is one of the 32 tribes who live in Jharkhand. This is a weaver artisan tribe. According to the survey of 1981 and 1991 their population amounts to about 40 thousands. It is very difficult to understand why their population is stable. Their central residing place is Ranchi, Gumla and Simdega districts. Their mother tongue is Nagpuri (Sadri, Sadani). Chik Baraiks had an important position among the 30 tribes residing in Bihar; but their population average was only 0.69% of the total tribal population. Apart from Jharkhand, Chik Baraiks are also in great numbers in Chhattisgarh, West Bengal, Orissa, Assam and Bangladesh.

10 Concentration

In different tribal inhabited villages, this community resides as an inseparable organ of the village. Though they are scarcely populated, there is no village in this region which does not have the members from this community. In each village there are only 10 to 15 families of this community excluding some exceptions of 25 to 30 families in a village. In such villages this community lives congruously between the chief tribes, namely Munda, Oraon & Kharia. In such cases they become linked with the culture and customs of these chief tribes. In the villages there is also a sufficient population of Thakur, Brahmans, Kshatryia, Kumhar, Bania, Teli, Lohra, etc. All these people depend on each other on the basis of functional relations. There are some villages where they are only Chik Baraiks in the village. There is a colony of about 120 Chik Baraik families in Siliguri (West Bengal).

11 Oraon or Kurukh

It is said that they ruled a place called Rohtasgarh (now known as Rohtas), but after the Muslim invasion they were forced to move to Jharkhand and were given shelter by the Mundas and the Sadans. Earlier they were known as Kurukh and when they came to Jharkhand they were named as Oraons by the Sadans and the Mundas. In due course they adapted Nagpuri as their means of communication. Surprisingly a book compiled by John Lakra (named "Mausmi Raag") contains 75% of the Kurukh/Oraon folk songs in Nagpuri/Sadri. Though when asked about their MT, they say Kurukh and not Nagpuri or Sadri.

12 Concentration

Oraon is the second largest tribal community of Jharkhand. They are scattered in the Chotanagpur province. Along with other tribes these people also migrated to Assam, West Bengal, Andaman, Orissa, Chhattisgarh and Bihar. But their main concentration is in Gumla, Lohardaga, Ranchi and Latehar Districts.

13 Nagpuri, Sadani or Sadri

It is said that to know anybody's culture, one must know his or her Mother tongue. From their mother tongue it is possible to know the ancient culture and beliefs of this region. Nagpuri is the pidgin of Jharkhand. In this state many ancient tribes like Munda, Oraon, Khadia, ChikBaraik and a total of thirty two tribes dwell together peacefully, all from different language groups (Austro Asiatic, Dravidian and Indo Aryan). The Mother tongue of Chik Baraik is known by several names like Sadri, Sadani, Ganwari, Nagpuri, Nagpuria and even Jharkhandi language in different parts of the country. "*Sadri is spoken over a vast area stretching from the eastern part of Madhya Pradesh to the Twenty Four Parganas in West Bengal and the Tea Gardens of Jalpaiguri to the borders of*

Orissa. It is known by many names such as Gawari, Sadani, Sadri, Nagpuria, Nagpuri, etc. It is used by a large section of the tribal as well as the non-tribal population either as a Mother tongue or as a lingua franca. Sadani / Sadri is a lingua franca used for inter-tribal communication in the eastern-central India. The term "Sadani" can be used in two senses. In the first, more general meaning, "Sadani" refers to the closely related linguistic varieties in Jharkhand, including forms such as Panch Parganiya, Khortha and Kurmali, which are generally considered independent languages. There are different views regarding the origin and status of Sadri. According to the Encyclopedia Mundarica, Sadri is the language of Sadans." "*Almost all the tribal people speak Sadani with some local variation*" Now also when villagers (I; e Munda, Oraon, Kharia, Chik Baraik, etc.) sit together, they use Nagpuri as a mode of communication. Nagpuri is used as the Mother tongue among the families of Munda, Oraon as well as Kharias residing in some parts like Simdega, Gumla, Palamu, Hazaribagh and Latehar Districts of Jharkhand who don't know their own Language. Same situation prevails in some parts of Orissa, West Bengal, Assam and Chhattisgarh, where these communities reside.

14 Language solidarity

It is noticed that the Chik Baraiks show their solidarity by speaking the pure variety of Nagpuri among themselves in the family. They have a strong affinity towards their mother tongue. They speak the second language or the other variety of the language when they come across non-Nagpuri speaker. They are well aware of the other varieties spoken by non Chik Baraik. The following are the prominent and easily noticeable distinct pronunciation that is made by the non Chik Baraiks.

English	Chik baraik	Oraon
He had come.	u ai rehe	u ai r he
They had come	u m n ai r h	u m n ai r h l
He will come.	ugo avi	u at u
They will come.	u m n ab	u m n abthun

15 Special quality of Language acquisition

Jharkhand is a multilingual state. Here different language communities co-exist. Due to this there is a lot of reciprocal influences. A lot of words are borrowed and exchanged. However it is remarkable that the Chik Baraik tribe has been able to keep the purity and the originality of the language intact. This is due to the special gift of language acquisition to this group. This group is known for acquiring the languages of the region in their accent, intonation and they can speak other languages (Mundari, Kharia, Kurukh, etc) like that of the native speaker. When a Chik Baraik speaks Mundari, Kharia, Kurukh or any other language of the region the listener would not identify that the speaker is a Chik Baraik. For example there is a Chik Baraik of Khunti District who is a very famous Munda folk singer and has been called several times to perform at the Ranchi Radio station and surprisingly he is a very good Nagpuri speaker as well. It is a known fact that the Chik Baraiks have been endowed with special ability of acquiring languages.

16 The use of Language

There is a special phenomenon among the Chik Baraiks that they use Nagpuri within the family set up. When they interact with other speech communities they speak Nagpuri but they would speak the language variety of the person with whom they are talking. That is, if a Chik Baraik is talking to

a Mundari speaking person, he/she would talk the Nagpuri of the Mundari variety with their tone and accent. If he/she is interacting with a Rautiya he/she would talk the variety of a Rautiya and the same with Oraon.

17 Prestige and low prestige varieties

In the case of Nagpuri there is no question of prestigious and low prestigious variety. Instead with regards to Nagpuri, people are more concerned and talk about its authenticity and purity, in its sound and phoneme. Purity in its use that is when one speaks Nagpuri there is no code mixing and code switching. One often comes across educated Nagpuri speaking people who openly accept that they are not able to speak the pure Nagpuri. A pure Nagpuri is spoken by an illiterate man and woman in the interior village. All accept this fact unanimously. An illiterate Chik Baraik woman would say "bais" [bais] 'sit down' and an educated Chik Baraik would say "baith" [baith] which is influenced by Hindi. The woman would say "muda" [muða] for 'but' and the educated 'lekin'. Thus the educated people show their inadequacy that they cannot speak the original and the authentic Nagpuri due to the influence of Hindi, English and due to their formal education. By speaking the original form of Nagpuri, a person's social status is not revealed instead people have high admiration for such a person who can speak Nagpuri with its authenticity and purity.

18 Difference according to class

Nagpuri speakers could be divided into literate and illiterate, city dwellers and village dwellers. The city dweller Chik Baraiks are very much influenced by the dominant language of city. Therefore it certainly has the influence over their Nagpuri language. For instance in Ranchi the people speak a variety which is very different from the so called original Nagpuri that is spoken in the district of Simdega. Those who dwell in the village speak pure Nagpuri. Those who are literate, as it is mentioned above, there is a lot of code mixing and code switching. Those who are illiterate they are well versed in Nagpuri vocabulary and they do not mix words from other tribal languages in contact.

19 Turn taking

In a normal conversation the listener follows the cooperative principle. They would not take their turn in the conversation unless they disagree with the person. Otherwise they would show their agreement by saying 'hou' [h u] means yes after each sentence, or equivalent response. If they disagree they would say 'nihi' [nihi] means 'no' and make their statement and give the reasons of disagreement. They end the conversation by saying 'lage tab jawa' [lage tab jawa] literally means, 'ok then you go' meaning you go ahead to your next intended work. And they disperse. While talking they do not use terms to confirm whether one is agreeing or disagreeing, as it is in English i.e isn't it, 'am I right,' etc. It is interesting to know that the Chik Baraiks do not express their gratitude explicitly. For instance when a guest comes he/she receives hospitality and while departing the guest would not say thanks explicitly but would invite the person to his/her house 'hamar bate aba tab' [hamar bate ba t b] 'you should come our side'. It implies that you come to our place too and give us a chance to serve you in return. The person responses 'lage tab khisate jaba' [lage t b khisate j b] literally it means 'ok then go ahead grumbling'. But the implied meaning is that 'we are sorry that we could not serve you well. If there was any inconvenience in our service please forgive us'.

20 Politeness and Respect

There are situations where the Chik Baraiks have to be conscious of their salutation and the manner of conversation is extremely polite and they show respect. A son or daughter would address their father calling 'tonye' [toe] 'you' but a prince or princess would address their father (ie; King) as 'apne' [apne]. 'tonye' [toe] is used among the equals and mostly among the younger generation. Among the elderly equals, they address each other by calling 'raure' [raure] 'you' term to show respect. A servant also addresses the son with the word 'apne' to show the respect and the son would call the servant with tonye' [toe]. The word [toe] is used among equals as the Hindi word 'tum'' 'you' and for respectful address second person singular 'aap' [a:p] is used. In Nagpuri for respectful address third person singular is used.

You eat food.	Chik Baraiks use different word form for honorific 'you'
[toy bʰat kʰa]	For the equals and without special respect.
[rͻure bʰat kʰau]	For father in law, sister in law, son in law, daughter in law, and for stranger
[tohre bʰat kʰawa]	For maternal uncle, nephews and in plural form.
[apne bʰat kʰay]	For the king, landlords and respectful person.

From where are you?	
[toy kahã kͻr hekis]	For the equals and without special respect.
[rͻure kahã kͻr heki]	For father in law, sister in law, son in law, daughter in law, and for stranger
[tohre kahã kͻr heka]	For maternal uncle, nephews and in plural form.
[apne kahã kͻr heky]	For the king, landlords and respectful person.

Whereas Oraons don't use honorific forms. Even there are disparities found in the language spoken by the Oraons of Ranchi District and Gumla-Lohardaga. For eg.

Where are you going?	Chik-Baraik	Oraons of Gumla	Oraons of Ranchi
For the equals and without special respect.	[toy kane jathis]	[toy kane ja in]	[tu kidhir ja rahle h]
For father in law, sister in law, son in law, daughter in law, and for stranger	[rͻure kane ja i]	[toy kane ja in]	[tu kidhir ja rahle h]
For maternal uncle, nephews and in plural form.	[tohre kane ja a]	[toy kane ja in]	[tu kidhir ja rahle h]
For the king, landlords and respectful person.	[apne kane ja ay]	[toy kane ja in]	[tu kidhir ja rahle h]

21 Taboos

10

Among the Chik Baraik certain names are restricted. For instance a wife would not address her husband with his name. She would use the third person singular 'apne' [pne] 'he' while she is telling someone about Husband and r ure while calling her Husband. Husband too would not address his wife by name instead he would call her by saying 'e babu ker main' [e babu ker maɪ] 'o mother of the son'. Similarly the wife does not call the name of her husbands elder brother or of the father in laws. Among the Chik Baraik they do not utter the name of the deceased family member. They would say 'marlaha rahe khane' [m rlaha rahe khane] 'when the dead one was alive then'.

22 Conclusion

During the British regime research works were done for Nagpuri, but in today's situation it's been neglected, and why because it is difficult to understand. Foreign scholars started doing research works from 1896. The foreigners like Conrad Buckout compiled grammar as well as dictionary and folklores but after their death, it was not even printed. Some grammar texts can be collected from the states like Orissa and West Bengal also. In the Indian context it should be noted that the Nagpuri literature is found in Meghalaya, Assam, West Bengal, Orissa and Chhattisgarh. The translated Nagpuri Bible was printed in Meghalaya in 1908 by Fr. P.Ponnette, we can see the importance of Nagpuri in 1908 in Meghalaya and this book surprisingly is not found in Churches of Jharkhand from where the language actually belong to. It is very necessary to throw light upon this language and its folklores, as it is the link language of three language families of the world. It was the official language (RAJBHASHA) of Chotanagpur since the reign of Phani Mukut Rai (0064 A.D.) till the British invasion (i.e. for around 1800 years). Not less than 4 to 5 Crore people speak Nagpuri, but it is difficult to find the actual number of Nagpuri speaker, as the Government has not conducted a Language Survey after 1961.

Knowing the importance of Nagpuri, some literate people (not the native speakers) like Sharma, Tiwari, Goswami, etc. started writing books which started affecting the language badly, because they started mixing Bhojpuri, Magahi as well as Hindi because of the lack of Standardization. Apart from this, more than five thousand Nagpuri audiocassettes, hundreds of music videos and more than ten films have been released and the progress is going on. Few years back a film 'Chal to Gori' released from Assam shows the importance and value of Nagpuri in Assam. There is a tradition of several magazines also in Nagpuri language. It was also announced then that if and when Jharkhand state was formed, Sadari would be the Official language of the State. Jaipal Singh Munda had started the Jharkhand movement (revolt for separate State), and he used to give speeches in Nagpuri for the people despite being a Munda. He had even stated, that "*When Jharkhand will become a separate State, 'Nagpuri' will be the State Language,*" but his dream remains a dream today also. Probably the reason behind not paying attention to this language is because of Abraham George Greirson's BIG Mistake of calling the Mother Tongue of Chik Baraik's as a Dialect. Nagpuri should be given the status of tribal language, as it is used by more than ten tribes as their MT. It is to be noted that Nagpuri has evolved from Prakrit and Apbhramsh in the forests of Chotanagpur independently just like Bangla, Oriya, Bhojpuri, etc. that can be justified from the book- Nagpuri Sadani Boli Ka Vyakaran1965 by Fr. P.S.Navrangi.

References

Ga jhu, Giridh r R m ("Girir j"). 2003. *mara gomke. (jayp l s h).* Ranchi: N gpur Prac ri Sabh .

Grierson, Sir George Abraham. 1903. *Linguistic Survey of India. Vol. V: Indo-Aryan Family, Eastern Group, Pt. II: Specimens of the Bihari and Oriya Languages.* Calcutta.

Jordan-Horstmann, Monika.1969. *Sadani. A Bhojpuri Dialect Spoken in Chotanagpur.* Wiesbaden: Otto Harrassowitz.

Lewis, M. Paul (ed.), 2009. *Ethnologue: Languages of the World, Sixteenth edition.* Dallas, Texas: SIL Interantional. Online version: http://www.ethnologue.com/

Nowrangi, Peter Shanti. ca. 1956. *A Simple Sad n Grammar.* Ranchi: D.S.S. Book Depot.

Nowrangi, Peter Shanti, S.J. 1957. *A Sadani Reader, containing stories, dialogues and songs. Companion to "A Simple Sadani Reader."* Ranchi.

Pras d, R m. 1992. *n gpur lok-kath .* Ranchi: Janj t y Bh   Ak dam , Government of Bihar. Encyclopaedia Mundarica.

Navrangi P.S.--- Nagpuri Sadani Boli Ka Vyakaran1965

Narmadeshwar Prasad- Land and People of Tribal Bihar 1961 B.T.R.I.

Baraik B.R.---Chik Baraik Ka Sankshipt Itihas1988

Hasnain N. – Janjatiya Bharat 1990

Sociolinguistics- Series 7 CIIL Mysore.

John Lakra- Mausami Raag.

Dr. Veerottam B.—Jharkhand; Itihas Evam Sanskrity2003

Dr. Ekka W. and Dr.Mishra S.S.—Chik Baraik Visheshank Dahar magazine. 1998

Naming Deficits in Bilingual Aphasia

Ridhima Batra   Pallavi Malik   Shyamala.K.C.

Department of Speech Language Pathology
All India Institute of Speech and Hearing, Naimisham Campus, Manasagangothri, Mysore
canif_ridhima@yahoo.co.in

Abstract

Naming calls into play multiple levels of processing. When used in conjunction with other basic tasks, it is usually possible to determine whether the principal cause of naming deficits are perceptual, semantic, or language output impairments. Naming of a visual stimulus such as an object or picture begins with early visual processing and recognition. The process of confrontation naming requires the formation of a perceptual representation of the object. It requires access to some sort of semantic representation to specify the concept that will then be tagged with the correct verbal label. Bilingualism is an intriguing phenomenon and has been defined variously by different authors. According to Fabbro (1999) people who speak and understand two or more languages and dialects are referred to as bi/multilinguals. Aphasia in a multilingual can lead to different language deficits in the languages known and is called as bi/multilingual aphasia. Naming deficits in aphasia are seen as retrieval failures which take different forms, depending upon the stage at which the breakdown occurs. A failure to retrieve the target lemma results either in selection of another lemma that has a similar semantic description i.e., semantic paraphasia; or a failure to retrieve a word's phonological description i.e., phonological paraphasia in which the word sounds like the correct word but sounds are substituted, added or rearranged; or a neologistic paraphasia with a production of a non-sense word. Paraphasias are common in aphasia and can help differentiate fluent from non-fluent aphasia. The aim here is to examine the naming deficits in bilingual aphasics and highlight the variation/correlation of these across languages. Six Kannada-English Bilingual aphasic subjects were taken for the study. All the participants were males with an age range of 25-50 years. The naming section of WAB, for all the subjects in both the languages was transcribed and analyzed for the presence of paraphasias (semantic, phonological paraphasias and neologisms). Variation of paraphasias among the fluent and non-fluent aphasic group and language specific errors in them were identified and described. The type of naming deficits exhibited by the fluent and non-fluent bilingual individuals with aphasia will be discussed in detail, in this paper.

1 Introduction

Naming calls into play multiple levels of processing. When used in conjunction with other basic tasks, it is usually possible to determine whether the principal cause of naming deficits are perceptual, semantic, or language output impairments. Naming of a visual stimulus such as an object or picture begins with early visual processing and recognition. The process of confrontation naming requires the formation of a perceptual representation of the object. It requires access to some sort of semantic representation to specify the concept that will then be tagged with the correct verbal label. Naming is a complex psychological function that can be disturbed in a variety of ways by the cortical and the sub cortical lesions.

Bilingualism is an intriguing phenomenon and has been defined variously by different authors. According to Fabbro (1990) people who speak and understand two or more languages and dialects are referred to as bi/multilinguals. Webster's dictionary (1961) defined a bilingual as having or using two languages especially as spoken with the fluency characteristics of a native speaker; a person using two languages habitually with control like that of a native speaker and bilingualism as the constant oral use of two languages. At one end is the approach of Bloomfield, (1933) who

defined bilingualism as "native-like control of two languages" whereas on the other end Haugen, (1950) takes a lax view by observing that bilingualism begins when the speaker of one language can produce complete meaningful utterances in the other languages.

Aphasia in a bilingual can lead to different language deficits in the languages known and is called as bi/multilingual aphasia. Aphasia in bilinguals can affect their languages equally or differentially. Bilingual aphasia has been a widely researched area as it provides insight into the brain functioning of a bilingual and the effect of a lesion on this functioning. Yiu and Worroll (1996) carried out investigations on a Cantonese – English speaker and reported that although the severity of aphasia was similar in both the languages according to standardized aphasia tests, a linguistic analysis of the data obtained showed that the pattern of disruption was not the same. The ability to construct sentences at the clause level and the use of morphological structures were relatively more disrupted in Cantonese.

One hundred years ago, Pitres (as cited in Paradis, 1995) reported that, contrary to expectation, some polyglot patients did not recover all of their languages to the same extent and pointed to a differential recovery pattern. He thought after a major clinical observation that the patient after his illness would retain the language he had been employing in daily life immediately before suffering from aphasia even though this may not be his mother tongue. This law came to be known as Pitre's law at a later stage. On the other hand, Ribot (as cited in Paradis, 1995) claimed that the language learned early in life by the patient was the last to be lost. This came to be known as Ribot's law and was at first an attractive hypothesis on neurological grounds in view of the vulnerability of the recent rather than remote memory to cerebral insults, and psychologically since it seemed to emphasize the importance of childhood training. These two rules have been cited in the literature repeatedly to stress the importance of factors such as automaticity of language learnt and usage of language as contributing to language recovery in aphasics. Thus, "Ribot's law" was cited as "primacy rule" and "Pitres law" came to be known as "familiarity (frequency) principle".

Naming deficits in bilingual aphasia are seen as retrieval failures which take different forms, depending upon the stage at which the breakdown occurs. It can be manifested in the form of paraphasic errors, perseveratory errors, circumlocutions or code switching errors.

Paraphasic errors-

The term "paraphasia" is applied to any unintended error of word or sound choice. Paraphasia is a symptom of commission that is an incorrect word is substituted for an intended or targeted word. It is the product of a breakdown at a stage of the word-retrieval process and, as such, is a dominant symptom within the more general category of anomia and is produced unintentionally. A failure to retrieve the target lemma results either in selection of another lemma that has a similar semantic description i.e., semantic paraphasia; or a failure to retrieve a word's phonological description i.e., phonological paraphasia in which the word sounds like the correct word but sounds are substituted, added or rearranged; or a neologistic paraphasia with a production of a non-sense word. Paraphasias are common in aphasia and can help differentiate fluent from non-fluent aphasia.

Paraphasias are common in aphasia and can help differentiate fluent from non-fluent output. Although phonemic substitutions do occur in non-fluent aphasia (Blumstein, 1973), they appear in a substrate of poorly articulated output and often represent dysarthric misproduction. The poorly articulated substitutions of non-fluent aphasia contrast with the substitutions of well produced but incorrect language components of fluent aphasia. Although some fluent aphasics may be aware of some of their paraphasia, most remain unaware of most of their substitutions.

Perseverations

The inappropriate repetition of a response to an earlier stimulus is a common phenomenon in many domains of performance by brain damage patients. Perseveration errors in naming are particularly common in aphasic patients. Perseverative reactions may inhibit the correct response by being stronger competitors because of their recency than a correct target that is difficult to retrieve.

Circumlocutions
Many patients compensate for a word retrieval failure by telling something about the object, in lieu of naming it. Aphasic patients may attempt to use such circumlocutions as self prompting devices, much as a normal speaker in a tip-of-the-tongue state may verbalize related, but incorrect words, in the hope of being led to the desired response.

Code switching errors
Code switching involves a complete shift to another language for a word, a phrase, a sentence and an utterance, or borrowing a word from the other language and integrating it into the base language. Valdes-Fallis, (1978) reiterated that in code switching, there is a clear recognition of each language in pronunciation and form. According to Milroy and Muysken (1995), Code switching is the alternative use by bilinguals, of two or more languages in the same conversation and requires a great deal of linguistic competence. According to Bhatia and Ritchie (1996), code switching refers to mixing of the various linguistic units (words, phrases, clauses and sentences) primarily from two participating grammatical systems across sentence boundaries within a speech event.

The increased use of switching between languages by aphasics may also denote linguistic deficits in both languages and the need to rely on available language codes to communicate more effectively (Krupa, Chegappa and Bhat, 2004). Munoz, Marquardt and Copeland (1999) compared discourse samples from four aphasic and four neurologically normal Hispanic Bilinguals in Monolingual English, Monolingual Spanish and Bilingual contexts, finding consistent matching of the language context by both the aphasic and normal subjects. However, the aphasic subjects demonstrated increased code switching, not evident in the speech samples of the normal subjects. This indicated increased dependence on both languages for communication subsequent to neurological impairment. According to Green (1986), aphasics possess weakened language systems in which executive processes in charge of inhibiting or activating one language over the other failed to do so effectively, thus resulting in more code switching errors. Fabbro (2001) reported that persistent mixing of elements from languages is associated with fluent aphasics and tends to be correlated with post-rolandic lesions whereas pathological switching between languages is associated with lesions of the frontal lobe (left and right).

Recovery patterns in the bilingual aphasics have been investigated by many researchers in the past. One of the earliest works in this area of research was conducted by Minkowski (1927) who reported a better recovery of the second language in almost 1/3 of the bilingual aphasics. He also identified the reasons why many polyglot aphasics did not recover their mother tongue and showed a better recovery of their second language. (1) visual factor, depending on the frequency with which the patient reads and writes in that language; (2) the affective factor, number of positive experiences related to the use of that language; (3) the environmental factor, namely the language spoken in the hospital setting; (4) the use of language in order to deal with specific topics which makes it highly automatized for some topics only; (5) conscious strategies applied during the acquisition process of a language; (6) linguistic factors, such as the degree of linguistic proximity of the two languages and (7) organic factors, example, the age of the patient and the type and extent of the lesion.

Aglioti and Fabbro (1993) studied the paradoxical recovery of the second language in a patient with a lesion to the basal ganglia of the left hemisphere. Recent studies on memory contribute to the

understanding of some aspects of paradoxical recovery of languages. For example, because the mother tongue is acquired unconsciously and at an informal level, it is mainly stored in implicit memory systems, whereas, the languages learnt through rules and used not automatically are mainly stored in explicit memory systems. Because aphasia mainly affects implicit memory systems, it is not uncommon to find patients who paradoxically tend to speak a language that they had never used before for communicative purposes more easily.

Paradis (1995) identified six recovery patterns. (1) Parallel recovery; (2) Differential recovery; (3) Selective recovery; (4) Antagonistic recovery; (5) Successive recovery; (6) Mixed patterns. But, this finding lacks support in Indian context.

2 Need of the study

The percentage of bilingual aphasics in India would be more, as the prevalence of bi/multilingualism in India is more. Hence studies relating to language of exposure are crucial and have both clinical as well as research implications in the bi/multilingual context of India, especially with increasing globalization. The present study is one such attempt and explores the nature of naming deficits in bi/multilingual aphasics.

3 Aim

To examine naming deficits in bilingual aphasics and highlight the variation/correlation of these across languages.

4 Methodology

Subjects

6 Kannada-English Bilingual aphasic subjects were taken for the study. All the participants were males.

Subject selection criteria

- Age range: 25-50 years.
- Subjects had Kannada as their mother tongue and had learnt English as second language before the age of 10 years.
- All the aphasic subjects had a history of left hemisphere cerebrovascular accident (CVA) confirmed by neurological examination and computerized tomography.
- The aphasic subjects were administered Western Aphasia Battery (WAB, Kertesz & Poole, 1974) for the identification of the aphasia type in subjects in both Kannada and English.
- Types of aphasia: Both fluent and non-fluent aphasic syndromes were considered which was decided on the basis of clinical observation and WAB-K (Kertesz & Poole, 1982) findings. Three fluent (2 anomic and 1 Wernicke's) and three non-fluent (2 Broca's and 1 Transcortical motor) aphasics were taken for the study.
- All subjects were right handed which was determined using self-report and information from significant others.
- Subjects with any auditory or visual deficit were excluded from the study.
- Ethical considerations were met.

Procedure

Subjects were seated comfortably. Then by casual talking, the subjects were made to feel at ease and the procedure was explained before the evaluation and recording began. The environment was made as distraction free as possible by carrying out the procedure in a quiet room and by the removal of any potential visual distracters. The entire verbal interaction with the subjects was audio recorded using Wavesurfer 6.0 software.

Test administered

Naming section of Western Aphasia Battery (Kertesz & Poole, 1982) was administered in both Kannada and English language.

Analysis

The naming section of WAB, for all the subjects in both the languages were transcribed and analyzed both quantitatively (statistical) and qualitatively (descriptive). The two groups (fluent and non-fluent aphasics) were compared for their performance in confrontation naming task using statistical measures. The comparison was also made across and within the groups. Also, comparison of performance in Kannada (L1) and English (L2) was made for both the groups.

The data was further analyzed for the responses without any cues and in the presence of semantic or phonemic cues, and tabulated. Also, categorization of responses in terms of error types like paraphasias (phonemic, semantic and neologistic), circumlocutions, perseverations, code switching, half words, gestural responses and no response (Shanatala, 1997), was also carried out to get an insight into the nature of the predominant errors seen in aphasics.

5 Results and discussion

The results are tabulated and presented with their quantitative and qualitative description on the following sections:
> Comparison of fluent vs. non-fluent groups
> Comparison of the aphasic subjects for L1 vs. L2
> Interference effects of L1 and L2
> Facilitation of responses with semantic vs. phonemic cues
> Categorization of error patterns

Comparison of fluent vs. non-fluent groups

Parameters	Aphasic groups	Mean	SD
Total correct scores (L1)	Fluent	10.3333	2.30940
	Non-fluent	3.6667	3.21455
Total correct scores (L2)	Fluent	16.3333	2.08167
	Non-fluent	6.6667	5.03322
Correct responses without cues (L1)	Fluent	6.6667	1.15470
	Non-fluent	.3333	.57735
Correct responses without cues (L2)	Fluent	14.0000	2.64575
	Non-fluent	1.6667	2.88675

Table 1: Means and Standard Deviations of total correct responses and correct responses without cues in L1 and L2 in fluent and non-fluent groups.

The mean and standard deviations for the two groups, namely fluent and non-fluent groups was calculated for correct responses with and without cues. It was found that the correct responses with

and without cues was better for the fluent group compared to the non-fluent group which was confirmed by the Mann-Whitney U test which showed a significant difference for both the languages (L1 and L2) (p<0.05).

Comparison of the aphasic subjects for L1 vs. L2

Parameters	Languages	Mean	SD
Total correct scores	L1	7.0000	4.42719
	L2	11.5000	6.31664
Correct responses without cues	L1	3.5000	3.56371
	L2	7.8333	7.19491

Table 2: Means and Standard Deviations of the total correct responses and correct responses without any cues for L1 and L2.

As is evident from the above table, all the aphasic subjects performed better in L2 (English) as compared to L1 (Mother tongue). This difference was found to be statistically significant on applying Wilcoxon Signed Rank test (p= 0.043). This finding draws support from the literature with respect to Pitre's Law which shows that premorbid language use plays a bigger role than the mother tongue. Also, as reported by Aglioti and Fabbro (1993), there is a paradoxical recovery of the second language compared to the mother tongue. Minkowski (1927) also reported better recovery of second language in bilingual aphasics probably because of premorbid frequent usage. This may have an implication on the importance of L2 being better than L1 depending on the variables like education, employment, issue of the language of prestige; rampant in the middle and higher class strata of the society.

Interference effects of L1 and L2

SUBJECTS	L1 → L2	L2 → L1
FA 1	2	11
FA 2	1	7
FA 3	0	10
NFA 1	0	2
NFA 2	1	2
NFA 3	4	3

Table 3: Raw scores of the interference effects of L1 on L2 and L2 on L1 for all the subjects.

By comparing the type of errors exhibited by the two groups in Kannada and English, it was noted that L2 interference on L1 was more for the fluent group as compared with the non-fluent group. This pattern was found co-occurring with their lesser and greater awareness, respectively with respect to code mixing. Mann-Whitney U test also showed a significantly higher interference of L2 on L1 for the fluent group (p= 0.046). This finding suggests that the subjects had difficulty in selecting L1 and could not inhibit L2. This phenomenon of lexical mixing was encountered when they could not find a word. Owing to anomia, bilingual aphasics are likely to substitute it with the corresponding word in another language (Perecman, 1984). However, there was no statistical difference obtained for the interference of L1 on L2 between the two aphasic groups (p=0.822).

Site of lesion studies have reported that persistent mixing of elements from languages is associated with fluent aphasics and tends to be correlated with post rolandic lesions whereas pathological switching between languages is associated with lesions of the frontal lobe (left and right) (Fabbro, 2001). While the present study did not focus on specifics of brain damage, the well assessed fluent and non-fluent aphasic patterns confirm to this finding.

Facilitation of responses with semantic vs. phonemic cues

Aphasic groups	Languages	No. of correct responses	No. of correct responses with semantic cue	No. of correct responses with phonemic cue	Total no. of correct responses
FA 1	L1	6	1	2	9
	L2	16	0	2	18
FA 2	L1	8	2	3	13
	L2	11	3	0	14
FA 3	L1	6	3	0	9
	L2	15	1	1	17
NFA 1	L1	0	1	4	5
	L2	0	6	6	12
NFA 2	L1	0	0	0	0
	L2	0	0	2	2
NFA 3	L1	1	2	3	6
	L2	5	0	1	6

Table 4: No. of correct responses with and without cues in Kannada (L1) and English (L2)

From the above table, it is evident that the responses of both the aphasic groups were facilitated by the semantic and phonemic cues. However, statistical difference could not be obtained between the two types of cues using Wilcoxon Signed Rank test ($p = 0.408$). This implies that both semantic and phonemic cues equally facilitated the production of correct responses in both the aphasic groups for both the languages.

Categorization of error patterns
Subject-wise number of correct and incorrect responses has been tabulated below:

Aphasic groups	Languages	No. of correct responses	No. of incorrect responses
FA 1	L1	6	14
	L2	16	4
FA 2	L1	8	12
	L2	11	9
FA 3	L1	6	14
	L2	15	5
NFA 1	L1	0	20
	L2	0	20
NFA 2	L1	0	20
	L2	0	20
NFA 3	L1	1	19

	L2	5	15

Table 5: Total number of correct and incorrect responses for L1 and L2 for all the subjects.

As illustrated in the above table, the number of incorrect responses was noticeably higher than the number of correct responses for the non-fluent group for both the languages and for the fluent group only in L1. So, the incorrect responses were subjected for further qualitative analysis by categorizing the error responses into nine types. The error types for all the 6 subjects in the respective languages are tabulated below.

Aphasic groups	Phone-mic parap-hasia	Semantic parapha-sia	Neolo-gism	Circum-locution	Persev-eration	Code switching	Half word	Gestural response	No response
FA 1	0	3	0	0	0	11	0	0	0
FA 2	1	4	0	0	0	7	0	0	0
FA 3	0	4	0	0	0	10	0	0	0
NFA 1	2	4	2	0	9	2	1	0	0
NFA 2	3	0	2	0	0	2	4	4	5
NFA 3	2	7	2	1	4	3	0	0	0

Table 6: Categorization of error types in Kannada

The above table describes that the fluent group has shown maximum of code switching errors (interference of L2 in L1) followed by semantic errors. The non-fluent group showed more number of paraphasic errors including semantic, phonemic and neologistic paraphasias followed by few code switching errors and perseverations. There was a marked difference in the type of responses by one non-fluent subject like half word response, gestural response and no response which were not prominent features exhibited by other subjects.

Aphasic groups	Phonemic error	Semantic error	Neolo-gism	Circum-locution	Persev-eration	Code switching	Half word	Gestural response	No response
FA 1	0	2	0	0	0	2	0	0	0
FA 2	1	6	0	1	0	1	0	0	0
FA 3	1	3	0	0	1	0	0	0	0
NFA 1	3	5	12	0	0	0	0	0	0
NFA 2	5	0	1	1	0	1	1	2	9
NFA 3	0	9	0	2	0	4	0	0	0

Table 7: Categorization of error types in English

The total number of errors seen in L2 were comparatively less and the type of errors also differed. The fluent group showed more of semantic paraphasias followed by a few code switching errors.

20

Whereas, the non-fluent group showed a maximum number of neologistic paraphasias and semantic errors and a few phonemic errors.

6 Conclusion

Clinical assessment and intervention procedures need to be looking at bilingual issues invariably. No matter how much it is ignored, multilingualism is here to stay and its specific needs should be addressed in our clinical management of aphasia.

The present study was focused on comparing the performance of bilingual aphasics on confrontation naming task. Results of the study revealed significant difference in the performance of fluent and non-fluent group wherein the fluent group scored higher on the given naming task. The naming deficits seen in the two groups also differed. The fluent group exhibited more of semantic paraphasias and code switching errors whereas the errors of the non-fluent group were mostly neologistic followed by phonemic and semantic paraphasias. This finding throws light on the fact that the fluent aphasics lack awareness. So, the clinical strategies for management of aphasics should focus more on awareness. The study also investigated the performance of the aphasics across the two languages (mother tongue and second language). All the aphasic subjects performed better in L2 (English) as compared to L1 (Mother tongue) which strengthens the corroborative evidences in the literature, besides highlighting the impact of several other variables that need to be considered.

7 Limitations of the study

There were some variables like severity of aphasia types which were not controlled could have had an effect on the results of the present study. Also, generalization of the results is limited only to this type of sample owing to the small sample size. The premorbid language proficiency was also not assessed very specifically through a standardized language proficiency scale (although there are no suitable ones) but only through informal information from the significant others, the premorbid language usage by the subjects was known. Control of all these variables would have probably given different perspectives of the findings.

References

Aglioti, S. and Fabbro, F. 1993. Paradoxical selective recovery in sub-cortical aphasia. *Brain, 119*, 1551-64.

Bhatia, T.K. and Ritchie, W.C. 1996. Bilingual code mixing, universal grammar and second language acquisition. In W.C. Rithie and T.K. Bhatia (ed.) *Handbook of second language acquisition* (pp. 627-689). SanDiego: Academic Press.

Bloomfield, L. (1933). *Language.* New York: Holt, Rinehart and Winston.

Blumstein, S.E. (1973). *A Phonological Investigation of Aphasic Speech.* The Hague: Mouton.

Krupa, E.D., Chengappa, S. & Bhat, S. (2004). Language mixing in Malayalam-English bilingual aphasics. *Asia pacific journal of disability and rehabilitation, 20, 120*-25.

Yiu, M-L.E. and Worroll, L.E. 1996. Sentence production ability of a bilingual Cantonese/ English agrammatic speaker. *Aphasiology*, *10*, 505-522.

Fabbro, F. 1990. *The Neurolinguistics of Bilingualism- An Introduction.* London: Psychology Press.

Fabbro, F. 2001. The Bilingual brain: Cerebral representation of languages. *Brain and Language, 79*, 211-222.

Green, D.W. 1986. Control, activation and resourse: A framework and a model for cobtrol of speech in bilinguals. *Brain and Language, 27*, 210-23.

Hamers, J. and Blanc, M.H.A.(2000). *Bilinguality and bilingualism (Second Edition).* Cambridge: Cambridge University Press.

Haugen, E. 1950. The analysis of linguistic borrowings. *Language, 26*, 210-231.

Milroy, L. and Muysken, P.(1995. *One speaker, two languages. Cross disciplinary perspectives on code switching.* Cambridge: Cambridge University Press.

Minkowski, M. 1927. Klinischer Beitrag zur Aphasic bei Polyglotten, speziell im Hinblick aufs Schweizerdeutsche. *Archives Suisses de Neurologie et de Psychiatrie, 21*, 43-72. Translated in Paradis (1983), 205-232.

Munoz, M.L., Marquardt, T.P. and Copeland, G. 1999. A comparison of code switching patterns of aphasic and neurologically normal bilinguals speakers of English and Spanish. *Brain and Language, 66*, 249-74.

Paradis, M. 1995. *Aspects of bilingual aphasia.* Oxford: Pergamon press.

Perecman 1984. Spontaneous translation and language mixing in a polygot aphasic. *Brain and Language*, *23*, 43-63.

Shanatala 1997. *Naming deficits in aphasia.* Unpublished dissertation. University of Mysore.

Valdes-Fallis, G. 1978. *Language and education: theory and practice: code switching and the classroom teacher.* Arlington: Arlington Centre for Applied Linguistics.

Webster, A.M. 1961. *Webster's sixth new collegiate dictionary.* Massachusetts: G.C. Marriam Company publishers.

ELDP Data Collection: Some Baram Experiences[1]

Dubi Nanda Dhakal, Tej Ratna Kansakar, Yogendra Prasad Yadava,
Balaram Prasain, Krishna Prasad Chalise & Krishna Paudel
Central Department of Linguistics,
Tribhuvan University
NEPAL

Abstract

Collecting data for a specific purpose is not as easy as some people think. The conventional method of text collections may be improved by making the fieldwork planned and systematic. Native speakers' knowledge of a language is associated with their feelings and emotions. In order to obtain specific genres their memories need to be stirred and stimulated. The task of data collection becomes more challenging and even frustrating if a specific community does not use it in natural settings. Linguistic and Ethnographic Documentation of the Baram Language (LEDBL) team has been working since May 2007 for the documentation of the Baram language, a language of the Tibeto-Burman language family spoken in Western Nepal. As the language has less than fifty speakers who use it in restricted domains, we encountered several problems at different phases of the data collection. With a well-planned, systematic and combined fieldwork during these two years, we have been successful in collecting representative corpus for the documentation of the Baram language. In this paper we briefly describe the methodology of data collection and finally suggest some recommendations for those who intend to work on the endangered languages. We include some morphological features of Baram in appendix A.

1 Introduction to LEDBL

The Linguistic and Ethnographic Documentation of the Baram Language (LEDBL) has been supported by Endangered Languages Documentation Programme, School of Oriental and African Studies (SOAS), University of London. The recent census reports that there are 342 speakers of the Baram language (Yadava 2001: 84). However, our sociolinguistic survey shows that the number of active Baram speakers is below fifty. The main objectives of LEDBL are to:

- To develop a corpus of the primary linguistic data of the Bar m language from every possible communicative domains existing in its speech community
- To elicit and record Bar m native speakers' intuition about grammaticality/acceptability of Bar m utterances, cultural load of meanings (e.g. taboos), ambiguity, kinship terms and other taxonomies, numbers and measures, dialectal variations, etc.

To develop the resources such as lexicon, sketch grammar, ethnographic profile, orthography, and primer on the basis of the database mentioned above.

2 Preliminary sociolinguistic survey

---

[1]   This paper is related to the Linguistic and Ethnographic Documentation of the Baram Language (LEDBL) supported by Endangered Languages Documentation Programme (ELDP), School of Oriental and African Studies, University of London, and hosted by Central Department of Linguistics, Tribhuvan University, Nepal. This documentation programme is led by Prof. Dr. Tej Ratna Kansakar and Prof. Dr. Yogendra Prasad Yadava, Tribhuvan University, Nepal.

We began our fieldwork in November 2007. The settlements where we worked is rather scattered at a distance of a few kilometers. The community members are engaged in agriculture and domestic chores and their economic status is average as compared to the people who live in their closest vicinity. In addition to agriculture, they even go outside their villages seeking seasonal employment like constructing houses and so on. In our first trip we rechecked the glossary[2] and made some corrections. Additionally, we also obtained some new vocabulary associated with flora and fauna. In order to obtain the wordlists, we used the semantic domains of *Ontology*. We prepared comprehensive sociolinguistic questionnaires in order to obtain information on the sociolinguistic situation of the language and assess its endangerment. This also helped us identify the speakers of this language and to some extent assess the proficiency of the native speakers. The sociolinguistic setting of the language reveals that this is a seriously endangered language when it is assessed in terms of the criteria set by UNESCO (2003). There has been a complete stop in intergenerational transmission of the language and this indicates the vital relevance for its documentation.

3 Data collection

Keeping in mind the objectives of LEDBL as the focus, we tried to capture varied and representative corpus. Our purpose therefore is to capture as many communicative events as possible based on Himmelmann (22). The communicative events or the manner people communicate in their language may vary from one speech community to another. As Himmelmann notes, the events are cultural-specific and the linguists working on the documentation of a language should capture all possible speech genres. He proposes the parameters of communicative events based on spontaneity as mentioned in Table (1).
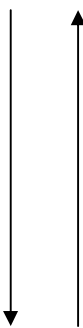
PARAMETER	MAJOR TYPES	EXAMPLES
UNPLANNED	Exclamative	'Ouch !' 'Fire !'
	Directive	'Scalpel !' greetings
	Conversational	chat discussion interview
	Monological	narrative description speech formal address
PLANNED	Ritual	litany

Table 1 Types of communicative events according to the parameter of spontaneity (Himmelmann 1998:22)

The speech events from 'planned' to/and 'unplanned' form a continuum rather than distinct antonymic pairs. Unplanned represents the pole of a continuum of spontaneity along which particular communicative events may be placed. Exclamation, for instance, is an unplanned communicative event. On the other hand, the rituals are performed after rehearsing several times

---

[2]   The glossary contains approximately 2200 words. Interestingly, the glossary contains the key vocabulary of the language. The financial support for the glossary was obtained from The National Foundation for Development of Indigenous Nationalities, Lalitpur, Nepal.

before it is performed. In a sense, these communicative events stand at poles apart in terms of 'planning of communication' and spontaneity.

Since our aim is to "*to develop the resources such as lexicon, sketch grammar, ethnographic profile, orthography, and primer*" we need to document a list of communicative events. At the outset, we were not sure whether there are any speakers who could speak fairly for a long time. We tried to expand the different topics that we would like to capture. The ethnographic topics based on Franchetto (189) were expanded and the whole cycle of ethnographic topics was listed. This included annual cycle of festivals, cultural rituals from menstruation to death. Capturing the ethnographic topics covered has multiple advantages: first, they cover the texts for preparing the ethnographic sketch, and second, they provide special kinds of texts for the language documentation. On a close observation, we noticed that the speech genres can be expanded by personal stories, reminiscences, and historical events and so on. Lupke's (77) inventory of speech genres helped us to identify the existing speech genres, and those which are indeed very rare in the Baram community. We adopted a rather general idea of genre rather than being very specific. We have followed Lupke (77) in identifying and making use of genres. He writes:

"In order to avoid identifying genres based on linguistic features rather than justifying them independently or using universal labels that lack a language-internal basis, I only distinguish genres in a very broad manner. Although I use existing labels as far as possible to identify them, these labels reflect my own intuitions and the culture-specific perspectives of Jalonke, except for clearly 'borrowed' genres" (77).

In addition to this, we consulted some grammars in order to plan the texts based on grammatical structures and topics. These grammatical topics were immensely useful to obtain specific language structures.

We expanded the genres and communicative events primarily for two reasons. Firstly, we are also making use of the genres in a very broad sense, whether this be a narrative or a travelogue, or to assign it any specific category, focusing mainly on the language documentation. Secondly, we have elicited various sessions aimed to elicit data for language use and structure. They mainly consist of several grammatical features like prohibition, conditional clauses, request, question and answer and so on. For example, we give them a hypothetical situation like, "If you got Rs. 50,000, what will you do?" and they would speak for some minutes. In addition to the texts used for the language documentation, they provide us specific language structures. We collected these sessions based on Leech and Svartvik (1995). These diverse communicative events, genres and grammatical sessions as a whole form our inventory of communicative events.

4 Inventory of communicative events

Baram is a seriously endangered language. Realizing this situation, we have expanded the communicative events proposed by Himmelmann (1998) and prepared an inventory of as mentioned in Table (1). This helps us capture the texts traditionally attested but also those related to diverse grammatical structures. Secondly, some speech genres need to be specific as they have several categories within a single genre. For instance, narrative is a universally common genre. This can further be categorized as 'historical narrative', 'personal narrative', 'reminiscences' and so on. Our inventory of speech genres, which is still being revised, is listed in the appendix B.

5 Speakers

Data of good quality depends upon the fluent language consultants of particular languages. We have considered ideal language speakers (Newell 1995:29), talk about good speakers of a language

to be chosen for a particular kind of work. However, in a speech community with a handful of language consultants, we did not have manifold options regarding the selection of the language consultants. At the beginning of the field work, we were not sure whether we would be fortunate to have fluent speakers who could speak considerably for a long time. The following paragraph presents the picture of a situation we passed through:

In the first week of our recordings we recorded the texts which were in fact very short. We were struggling to make good recordings. In our field office, one day, my friends Krishna Chalise and Krishna Paudel were with Mina Baram to record a text. I was out of our studio. As they came out of the room, they were so delighted that they were fortunate to have a fluent speaker who could speak very fluently for an hour. At that point we were very hopeful about collecting data. The days following this were full of excitements and we were very delighted to have diverse texts and representative corpus.

Grinevald (2003:64) presents the spectrum of language speakers and classifies the speakers of endangered languages into four categories (i.e. (a) Native fluent speakers, (b) Semi-speakers, (c) Terminal speakers, and (d) Rememberers.

Baram communities do not have the speakers of the first category. We do not have speakers who are monolingual in the language, and whose dominant language is Baram. We therefore have to rely on the semi-speakers. The language consultants we are working with are bilingual whose dominant language is Nepali although they are fluent in Baram. They do not use Baram regularly and naturally. One of the speakers commented that she was happy to talk to her sister in Baram after an interval of a few months. However, they are capable of producing 'the best texts' ever existed in the language, which otherwise is likely to go to oblivion.

We have the speakers of the third and the fourth categories as well, namely terminal speakers and rememberers. On the basis of their mastery of the language, and mastery of the different genres, we rely heavily on the texts produced by the speakers of the second category. We involved the speakers of the second category, but not the others except for a few sessions.The usefulness of semi-speakers in language documentation is mentioned by Grinevald (67):

They are generally essential to projects of documentation, because of their knowledge of the language and they being fully bilingual and generally the best trained people in community, even if it turns out that their language proficiency is more limited than professed when time comes to actually do linguistic analysis, Although old fluent speakers seem to be the obvious speakers to seek out, and ultimately are the source of the major recordings of the language, they may or may not make good linguistic informants, depending on their age and sophistication, and their analytical and linguistic talents.

These language consultants are the best contributors we have for the language documentation. They are all bilingual, learned Baram as their mother tongue, but now gradually shifted to Nepali as a language for day to day communication. Although they are 'semi-speakers', they possess the best features of the Baram language. Once these speakers pass away, the language will go to oblivion.

Among all the speakers only one language consultant has passed grade three, whereas the remaining language consultants are either prelliterate or hardly literate. This makes the task of transcribing and translating difficult, and elicitation of grammatical items and preparation of paradigms even more challenging. In our ELAN Annotated files we have transcribed the texts in IPA and Devanagari, and translated into Nepali and English. If we had got the language speakers who could transcribe in IPA and Devanagari and translate into Nepali, the total duration of our annotation files would have been dramatically increased. Till this time we have been struggling to transcribe and translate the texts we have so far collected.

The task of collecting data is followed by activities like transcribing, translating, and annotating in a prescribed format. Unless data of good quality is obtained, the later stages of analysis become

difficult. We now realize that the data of good quality will lead to faster transcription and translation, and thus serve to gear up the activities related to analysis. Collecting data is not an end task in itself but the beginning of a long chain of activities. The good recording, the richness of data and coherence in the texts will lead to faster transcription, translation, and all related activities.

The annotated lines are exported after ELAN-Toolbox conversion. These lines are transcribed in ELAN, and interlinearized in Toolbox. The 'ref' indicates the number of each chunk in a particular file, 'ut' is the utterance of the language consultants and ELAN Begin and ELAN End indicate the time each utterance begins and ends respectively. They are illustrated in (1).[3]

(1) \ref 001
    \ut b m na b ndisko
    \ELANBegin 00:00:01.020
    \ELANEnd 00:00:02.610
    \ELANParticipant TOK

    \txd अब मना बान्दस्का

    \tx b m na b ndisko
    \mb b m na b n -di -go
    \ge now   yeast make -NTVZ -INF
    \ft Now making yeast

    \ftn अब मचा बनाउने.

6 Methods

If you are working on an endangered language, it is difficult to obtain the kinds of data you try to capture without preparation. For example, all our speakers can contribute to some extent to the 'personal narratives' or 'reminiscences' but we find it difficult to obtain data on specific topics like 'prohibition' or 'hortative' and so on. It is even more difficult to get specific paradigms unless some necessary arrangements are made. The followings are some of the techniques we used during our fieldwork

6.1 Direct elicitation

This is the technique we use for preparing glossaries, making paradigms, and data for specific topic while writing the Baram grammar. Some speakers are capable of this whereas others are not. One speaker can tell us the paradigm very easily but may not be sure of some linguistic knowledge. The person who knows better cannot understand what we are trying to elicit. A group of language informants seated together and eliciting the glossaries and paradigms is extremely useful instead of working with a single speaker. Some features of languages are not well probed unless we work with multiple speakers. We hope to continue this process till the end of the project until a sketch grammar is written and a dictionary is prepared. We realize that a person who has some formal education is extremely useful for this kind of elicitation.

---

[3]   The abbreviation conventions are as follws: ABL- ablative; AUX -auxiliary; CAUS -causative; CLF-classifier; COND -conditional; DAT-dative; ERG- ergative; ft- free translation; ftn- free translation in English; ge- gloss in English; GEN-genitive; HAB -habitual; HAB -habitual; INF- infinitive; LOC -locative; LOC-locative; mb-morpheme break; NPST- non-past; NTVZ-nativizing suffix; PFV- perfect; PST-past; ref- reference; SEQ- sequential converb; SEQ-sequential converb; tx- text; txd- text in Devanagari; ut- utterance.

6.2 Narration

In a general sense everyone is a storyteller. All people tell stories and thus narration is a skill all people are familiar with. This is the main method we used for collecting large amount of texts. The language consultants tell about their personal narratives, or reminiscences, or travelogues. We tried to correct, advice, or orient them if the texts they produce deviate in content. No sooner we realize that the speakers deviate from the theme, we immediately stop them. We rehearsed somehow to make them tell the stories in the manner we intend. Sometimes, we tell them about describing incidents, telling personal reminiscences, and so on before they actually produce the texts. The speakers were trained when they worked with us for some weeks.

"Could you tell us about …."How wine is prepared?" or "How did you travel from Dandagaun to Gorkha?...." Several texts were obtained using this technique. Although we were not sure whether we could collect the data we intended, we were very successful incollecting data by using this technique.

6.3 Role play and simulation

A large number of speech genres can not be obtained simply by narration. We therefore tried to obtain the corpus by role play and simulation. The data related to specific grammatical items are difficult to obtain. In order to obtain these kinds of data, we advise the speakers to involve in role play/simulation. For example, we have a session where a speaker asks permission of his elder brother to visit his relatives. His brother replies that this is the season of farming, and he should not go now.

Using these techniques, we got specific grammatical items. Imperative sentences may be elicited but if they are obtained through role play and simulation we get repeated structures. If the same structures or constructions are repeated again and again, it is easy to be certain about specific grammatical constructions. We have several recordings obtained through role play and simulation. Several dialogues were recorded using this technique. We intend to use this method to construct specific grammatical structures. Similarly, there are neither riddles nor proverbs in Baram of their own! However, we asked them to tell us some and made recordings related to such expressions.

6.4 Stimuli used

It was necessary to use stimuli in the process of data collection. Some structures are related to certain kinds of activities, or objects. People can easily describe the things they see. For example, we found that pictures can be used to elicit structures like, 'This is ….on/in/back/forward' or 'There is a stone on ….' We had sessions related to postpositions, and adverbs. If one intends to obtain structures like these, or the use of postpositions, or general descriptive texts, the use of stimuli is very effective. We describe below the various ways of using stimuli in data collection.

(a) We showed them documentary films and later asked them what happened in the film. Several speakers described the film as they understood it.

(b) We showed them songs and later asked them what the song clips are about. They then describe what happened in it.

(c) We showed them the pictures of objects and asked them what they are used for, how they are prepared and so on. We had sessions like 'Use of Bamboo' by showing them the picture of bamboo and bamboo objects.

(d) Questions were shown and displayed on a laptop. One reads the questions (asks the question to the target language consultants and the target speaker gives the answer). They were very useful to

obtain certain kinds of structures, i.e. purposive clause, answer of 'wh-question', and so on. These kinds of structures are very specific.

7 Setting: Natural or controlled?

A majority of files were recorded in controlled settings in our field office in Gorkha. We do not generally prefer the natural setting due to the disturbances caused in the environment. Our recordings need to be free of outer disturbances and distraction. We have therefore decided to record in the field office rather than in natural settings. The hissing sounds, clicks, chirps of birds, and cries of animals, barking of dogs, etc. are some unwanted sounds we faced during our recordings. In order to avoid these sounds, we decided to record in the field office.

8 Equipment and Software

For recording the textswe used the audio and video recorder, and microphones. In order to make good quality recordings, we have also set up a studio for recording in our field office in Gorkha. The following are the computer softwares we are using:

1.   ELAN (EUDICO linguistic annotator is an annotation tool that allows to create, edit and search annotation for video and audio data. It was developed at the Max Planck Institute for Psycholinguisitcs, Nijmegen, the Netherlands. (Version 3.6.0).

2.   Toolbox (The Field Linguist's Toolbox is a computer program that helps field linguists and anthropologists integrate various kinds of text data: lexical, cultural, grammatical, etc. (Version 1.5.3).

3.   Audacity (for editing media files (Version 1.2.6)).

4.   Adobe Premiere Pro (for video editing (Version 7.0)).

5.   IMDI (for metadata editing (Version 3.1)).

9 Problems

During the annotating period we had a number of problems which can be listed as follows:

(a)   It is difficult to get fluent and cooperative speakers. Some language speakers are good speakers but are unwilling to work. There is an aged woman who could tell good texts but her articulation is not clear. She has her front teeth missing! Some speakers are useful but the language they retain is not very good.

(b)   Several texts have no coherence at all. This causes difficulties in translation and annotation. We realize that we are unable to say why such irrelevant utterances occur throughout the texts. We also need to notice that lack of coherence is related to particular language consultants.

(c)   Such texts also lack richness of linguistic features. In the discourse data the language seems to be merely the translation of the 'Nepali texts' into the Baram language rather than Baram texts *per se*. Therefore, the richness in terms of 'native features' and 'linguistic richness' should be assessed before we continue to record the texts.

(d)   Sometimes, fieldwork is hampered by equipments. We encountered several problems with the video recording. It is therefore a good idea to purchase good quality equipments like video camera, still camera, microphone and laptops for the fieldwork.

(e)   Since fieldwork is related to language documentation, and is also associated with the kinds of facilities they receive, the facilities given to the language informants should be clearly mentioned before they are employed. It is natural that the dealing with language consultants must be settled very fairly and honestly.

10 Our corpus

We were not sure about the kind of data we would gather in the field in the early hours. In the early hours of the days, the longest sessions we recorded lasted for a few minutes. Coincidently we had the first session with Mina Baram. She proved to be one of the best contributors for our texts in terms of linguistic features.

The work progressed and the recordings went on. We now have a total of about sixty eight hours of recording. The principle genres, and sessions mentioned in Appendix B have been covered. However, some sessions are very rare in the data. Some speech genres are still missing. Some formulaic expressions such as leave taking, taboos, urging, proverbs, etc are difficult to get in an artificial setting.

These topics were covered only during the last week of October and the first few weeks of November 2008 after systematic planning. For example, it is difficult to get a topic on 'announcing' because the language has not currently been used for announcing anything. Similarly, 'greetings' are not exchanged in Baram when they meet. We did not have any way to capture such sessions. We simply ask them to describe or report what they used to do in their traditional greetings and leave taking. Although this does not directly describe the greetings and leave taking, this reminded us of greetings in the community. Some formulaic expressions like proverbs, ritual wailings and taboos were difficult to capture because of their restricted use.

11 Representative corpus?

With the detailed planning, preparation and strategies, we have tried to make our data representative in terms of genres, communicative events and sessions related to grammatical structures. Some genres are more common compared to others because narrative is found in all communities, but particular formulaic expressions like greetings and leave taking may not exist in a language like Baram. We have detailed descriptions on birth and death rituals but we failed to capture a session like ritual wailing. With all these methodologies, we have done our utmost to make our database more representative.

12 Conclusion and recommendation

All the texts that we have collected are extemporaneous and oral texts. Our corpus basically lacks very formal genres like speeches, announcements, and so on although we have some texts obtained through simulation. Collection of quality data is really a challenging task when we work with minority and endangered languages. It is even more difficult if the speakers do not use the language in natural setting. In our case, all data we obtained were by the speakers who have stopped speaking their language in natural settings. Although we became desperate in the early phase of our field work, we finally recorded and made the data more representative. Sometimes we were happy to record the sessions which ran for a long period of time but the quality of linguistic contents in such sessions was not rich. An analysis of the linguistic features of each speaker leads us to further confirmation that the speakers that we have are worth working with. We like to conclude this paper with a few suggestions for the field linguists:

1. Make an inventory of different genres, communicative events and ethnographic topics before beginning the fieldwork. This helps you capture diverse sessions with different uses of the language. The objectives of the research also determine the method of collecting data.

2. Work with different speakers. One may be good at narrating stories, another at giving instruction and another perhaps at data elicitation for making paradigms. As a field worker, we need to be aware of the capability of different speakers.

3. Make a proper analysis of the data that you have collected before you record texts for hours! You will be surprised to see that the linguistic contents of the speaker who speaks for a long time is perhaps worthless! A sample analysis of each informant you are working with will suffice to determine whether the text one contributes has coherence, unnecessary repetitions, hesitation markers, unusual pauses, unusual breathing and so on. We find that one of the speakers we worked with always clears his throat at the interval of a few sentences!

4. Make your sessions of moderate length. If they are very long, i.e. for an hour, the data may be difficult to handle in some computer software like ELAN, TOOLBOX, AUDACITY when these files are to be edited and annotated. In our long files, the speakers deviate from the topics they speak on. If your sessions are very short, there may be innumerable files but they can neither increase the size of the corpus nor the total duration.

5. Record some sessions which are directly related to the grammar, i.e. 'conditional clause', 'purposive clause', etc. This will be a better way of obtaining a specific structure rather than to ask them to translate the constructed texts.

6. The language speakers should be properly trained and oriented while recording these specific topics.

Dealing with the language consultants is a key factor during the data collection period if a documentation project runs for a few years. They should be honestly and fairly treated.

Appendix A. Morphological features

The appendix highlights some of the morphosyntactic features of Baram. Table (2) gives the outline of Baram pronouns and their inflection for different cases.

PRONOUNS		CASES			
		ERG	DAT	GEN	ABL
1SG	a	i	ag i	ago	aga
1PL	Ni	Nie	nig i	nigo	niga
2SG	Na	n e	na g i	na go	na ga
2PL	Nu	Nue	nu g i	nu go	nu ga
3SG	u	us i	usg i	usko	usga
3PL	uba	ub	uba g i	uba go	uba ga

Table 2 Baram pronouns

Only three numerals, the numbers one to three exist in Baram, i.e. *de* 'one' *nis* 'two', *som* 'three'. Classifiers attach directly to the noun stem preceding the nouns. Only two native classifiers *-wa* and *-e* are found in Baram. They follow the animate and inanimate nouns respectively. The classifier -*wa* 'CLF' comes only with the numeral *de* 'one' *nis* 'two' and *som* 'three'. All other numerals are borrowed from Nepali. When the classifier *-wa* 'CLF' occurs with the numeral *de* 'one' its form is *dzewa* 'one-CLF'. They can not be separated. However, when it occurs with the numeral *nis* 'two' or *som* 'three' its form is distinct as in (2)

(2) a. *tjo de dz na kjajo*
         tjo   de   dz na   ki-ajo   i- a
         that one   CLF   PST-look
         'A man grazed the cows.'

b.   *hai so dz na huku hai hai kita ni a*

    hai   som   dz na huk-u   hai   hai   ki-ta

    what   three   CLF   sit-SEQ   what   what   PFV-keepat

    ni- a

    NPST-AUX.NPST

    'What have the three of them done ? '

c.   *dzewae s ppei cisja a*

    dzewa-e   s ppei   ci-sja -a

    one.CLF-ERG   all   HAB-drink-PFV

    'One would eat all the things.'

In addition to a very few lexical adjectives, a large number of adjectives are derived from verbs, *uchogo* 'to become fat' *kjocho* 'fat', *osungo* 'to become cold' *kjosun* 'cold', *ohongo* 'to become hot', *kjohon* 'hot', *uogo* 'to become yellow', *keuuo* 'yellow'. Adjectives occur attributively and predicatively (3).

(3) a.   *alam dzat ikini dzat*

         alam   dzat   ikini   dzat

         big   caste small caste

         'High caste, low caste.'

    b.   *gjabo mu th jo*

         gjabo mu   th ja-o

         white clothes   bring-IMP

         'Bring the white clothes.'

    c.   *ikinse hãga laku gidzen*

         ikinse hãga   laku   gi-dzen

         small   branch   cut-SEQ   PST-split

         'A small branch (of a tree) was cut and splitted.'

Some of the Baram affixes are prefixes whereas the others are suffixes. Infinitive form of a Baram verb is *<ko-/go->* which occur with the voiceless or voiced stem respectively. Past tense markers are *<ki-/gi->*. When a verb from Nepali (or Indo-Aryan languages) is used in Baram a nativizing suffix *-di* is added before the infinitive marker is added. For example, *tar-* 'to take across' occurs in the Baram language, the nativizing suffix *<-di>* is suffixed to the verb. The borrowed stem thus becomes *tardigo* 'to take across' from the borrowed stem *tar-* 'to take across'. We have thus the stems *t un-di-go* 'to imprison' or *re -di-go* 'to cut the throat' as the verb stems are borrowed from the Nepali language. Non-past tense marker is *i-*.

(4) a.   *m uri asi go d i kja*

         m uri   asi -go   da-i   ki-ja

         bee   take-INF   say-SEQ   PST-walk

         '(They) went to take out the honeybees.'

    b.   *am m kica*

         am   m   ki-ca

         rice   also   PST-eat

         '(He) also ate rice.'

(5) a.   *tjo sjãku lakko nikham*

         tjo   sjãku   lak-ko   ni-kham

         that   saal   cut-INF   NPST-say

         'They say to cut down the saal trees.'

b.　　*kumba niceplo*
　　　　kumba　　　ni-ce-uplo
　　　　stone　　　　NPST-CAUS-be turned over
　　　　'They make them take out stones.'

Progressive is encoded with *<-i >,* perfect is the same as the past tense marker, habitual markers are *<ci-/dzi->.* Conditional is coded by the suffix *<-le >,* desiderative marker is *<-se >* Imperative is *<-o >* andative (direction away from the deictic centre) is *<he- >,* and venitive (direction towards the deictic centre) is *< he - >.*

(6)　a.　　*a chanag i hja i cihuka*
　　　　　 a chana-g i　hja -i　　　　ci-huk-a
　　　　　roof-LOC　　climb up-SEQ HAB-eat-PFV
　　　　　'I stayed by climbing to the roof.'

　　　b.　　*ibi case ni a*
　　　　　ibi　　ca-se　　　　ni a
　　　　　REFL　eat-DESID　　AUX.NPST
　　　　　'I want to eat.'

　　　c.　　*asa cuno*
　　　　　asa　　cun-o
　　　　　oil　　add-IMP
　　　　　'Add some oil.'

　　　d.　　*sja n nu mahale han u g i ada p na*
　　　　　sja　　n nu　maha-le　　　　　han u -g i　　ada　　　p na
　　　　　cow　　milk　NEG.COP-COND　　village-LOC　search　should
　　　　　'If there is no cow milk, one should search in the village.'

Appendix B. Inventories of communicative genre for text collection

S N	Communicative Events	Text type/Genres
1	Exclamative	Cries, Signs of surprise, joys, etc.
2	Directive	Prohibition, suggestion, instruction, recommendations, urging, ordering, permission, telling (somebody to do something etc.), vocative
3	Conversational	Conversation, chat, discussion, interview, plays, asking (about activity, trouble/problems, thoughts/feelings, health/physical states, etc.), requests, suggesting/advising, apologizing, songs, procedural texts, reporting, clarifying, debate, accusing
4	Monological	Narratives (personal, historical, myth, travelogue, reminiscences), speech, routines, announcing, illustration with examples, comments, opinions (likes and dislikes, feelings), quoting other people, reported statements and questions, proverbs, songs, procedural texts, action description, reporting, clarifying, debate, description
5	Ritual	Rituals (birth, naming ceremony, marriage, ritual wailing), festivals, ghosts, first menses, folk dances, cultural observations
6	Others	Letter, play, description (action, place etc), turn taking rules, formulaic expressions (proverbs, greetings, leave taking), language play (riddles, humor), taboo, supernatural

References

Abbi, Anvita. 2001. *A Manual of Linguistic Fieldwork and Structures of Indian Languages.* Lincolm Handbooks on Linguistics, No.17. Munchen: Lincom Europa.

Bouquizux, Luc and Jacqueline M.C. Thomas. 1992. *Studying and Describing Unwritten Languages.*Trans. James Roberts. Dallas: Summer Institute of Linguistics.

Franchetto, Bruna. 2006. "Ethnography in Language Documentation*." Essentials of Language Documentation*. Ed. Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel. Berlin and New York; Mouton de Gruyter, 183-212.

Grinvald, Colette. 2003. "Speakers and Documentation of Endangered Language." *Language Documentation and Description,* Vol. 1. Ed. Peter K. Austin. London: Endangered Language Project, 53-72.

Himmelmann, Nikolaus. 1998. "Documentary and Descriptive Linguistics." *Linguistics,* 36:161-95.

Leech, Geoffrey and Jan Svartvik. 1994. *A Communicative Grammar of English.* Singapore: ELBS.

Lupke, Fiederike. 2003. "Small is beautiful: contributions of field-based corpora to different linguistic disciplines, illustrated by Jalonke." *Language Documentation and Description,* Vol 3. Ed. Peter K. Austin. London: Endangered Language Project, 75-105.

Newell, Leonard E. 1995. *Handbook on Lexicography.* Manila: Linguistic Society of the Philippines.

Prasain, Balaram and Krishna P. Chalise. 2004. Eds. *A Basic Baram-Nepali-English Dictionary*. Kathmandu: The National Foundation for Development of Indigenous Nationalities.

Samarin, William J. 1967. *Field Linguistics: A Guide to Fieldwork.* New York: Holt Rinehart and Sinston, Inc.

UNESCO. 2009. "International Expert Meeting on UNESCO Programme: Safeguarding of Endangered Languages". www.unesco.org/culture/inh/doc.downloaded on 13 February.

Vaux, Bert and Justin Cooper. 2003. *Introduction to Linguistic Field Methods.* Munchen: Lincom Europa.

Yadava, Yogendra P. 2001. "Language". Population monograph of Nepal, i. Kathmandu: Central Bureau of Statistics and UNFPA, 137-172.

Skeuomorphism in P nini

Chinmay Vijay Dharurkar
Department of Humanities and Social Sciences,
Indian Institute of Technology, Bombay,
Mumbai, INDIA
chinmayvijay@gmail.com

Abstract

This is an attempt to bring a concept in the history of Design, Architecture or Archeology called Skeuomorphism to Linguistics or to be specific, Indian (Sanskrit!) Intellectual History. Besides elaborating and adapting it to the Sanskrit Grammatical tradition, certain concrete terms have been elaborated upon as how they epitomise Skeuomorphism. Skeuomorph refers to a derivative object which retains ornamental design cues to the structure that was necessary in the original. A feature in a design remains even after losing its function; this is at the heart of Skeuomorphism. It has been attempted here to show how certain terms in Sanskrit grammatical tradition are skeuomorphic. Three terms: vibhakti, puru a and sarvan ma have been discussed in detail and are shown to be skeuomorphs. More terms from the appendix A can be potential candidates to be skeuomorphs, as these are the terms that are found in pre-P inian grammars in form and still P ini defines in his formal grammar. And obviously P ini makes the concept formal and non-prototypical.

1 Skeuomorphism

Skeuomorphism is a term in the discipline of Design where it refers to a design-feature functionally void but appears in a design, simply because the users have been used to the specific feature. A common example quoted for a skeuomorph is that of the click sound that is featured in the modern-tech digital cameras. Here, the click sound is not an outcome of the very design of the modern-digital-technology-based digicams, as it used to be of the old analog cameras. So, it has no functional significance. It is in fact put into the cameras by an additional chip, which controls the click-sound. So, the click sound in modern digicams is a skeuomorph of the old functionally significant click sound.

Similarly, in any discourse or discipline, old terms keep gaining new, richer or deviated senses in the course of time. These terms are not born anew, most of them are potential candidates to be skeuomorphs of the older ones.

By putting it in terms of Skeuomorphism, we are not being fashionable nor are we trying to be prolix by re-instating the already known facts about the semiotics of nomenclature or terminology.

Like any other knowledge system, Sanskrit Grammatical Tradition being no exception is bound to have skeuomorphs. In fact, when we can put something in terms of tradition, it is bound to exhibit Skeuomorphism of some degree.

2 Why Skeuomorphism

It becomes important to see the technical terms in Sanskrit Grammar as skeuomorphs, as it focuses on the discursive deployment of the terms in the tradition, where P nini may be thought to be a comfortable and important point to study.

It is insightful to see the terms as skeuomorphs because it gives us a picture of the diachronic development of the knowledge base in the field of Sanskrit Grammar in general and P ninian tradition in particular.

It is significant as it opens a fertile area of the intellectual history of Sanskrit and P ninian grammatical tradition and comes through as an effective revealing tool in this intellectual historiography.

If we simply view a term as possessing multiple meanings i.e. being polysemous, then we may not necessarily be missing the diachronic development of the term but we may ignore the development of the term as *how* it has reached to the meaning what it has, at a particular point of time.

Skeuomorphism is helpful to see how the meaning of a term has been negotiated and changed over time, in the semiotic system of the technical terms and terminology.

It helps us contrast the functional apparatus set up by a term, whenever it is recycled and reused.

3 On the development of the nomenclature and terminology in P nini

3.1 How is P nini different from his predecessors

Technical terms in P nini are not the terms that have exclusively emerged only in P nini. There are certainly pre-P ninian elements in the A dhy y . So a lot of the pre-P ninian terms have been used with alterations in the grammar. This recycling and reuse of the terms has resulted in an alteration that in some cases, it is only the form of the term that has remained the same, but the formal implications of the terms[1] have been thoroughly revised.

The pre-P ninian grammars mainly are the Vedic pr ti khyas, which are concerned mainly with the phonology and recitation of the Vedas. Another source is the Nirukta, where we find a whole lot of pre-P ninian grammatical terms. We shall not go into the details of the pre-P ninian literature, as detailed treatises on it are already available.

One main reason for P nini's grammar becoming popular and being widely taught in the then traditional academia was its formal rigor. A clear-cut shift of the implications can be seen in the deployment of the terms that were traditionally used. P nini makes them formal. He tries to do away with the semantic implication and tends to provide a formal enumerative definition. This is very clear in case of the terms like vibhakti, pratyaya, dh tu, sarvan ma, etc.

P nini also develops a distinct abstraction in his descriptive technique. The way the concept of person has been optimized clearly betrays the rigor of the abstraction in P nini. For example, it is notable that P nini uses the older notion or concept of person but does not use the term *puru a,* technically. This very idea of reducing significant grammatical concepts to the (abstract) morphemic units is very typical of P nini. So also for P nini the abstract idea of case has been done away with, and the term vibhakti has been reduced to a set of the (abstract) morphemic units or suffixes.

Thus, though P nini is not detached from his predecessors, he has given an altogether different formal grammar of his own. He had developed explicitly distinct concepts, ideas and descriptive techniques that are not found in the earlier descriptions or grammars. So skeuomorphic terms are bound to be there in P nini. When it comes to the parallels in the pre-P ninian and the P ninian there are instances where P nini differs in the formal rigor and totally innovative descriptive technique.

---

[1] Please see the appendix A for a quick view of the terms that are pre-P inian in form and meaning and defined by P ini. It is out of these terms where we can enlist and study the skeuomorphic terms in P ini. It is here where we find terms that can be potential candidates to be skeuomorphs.

3.2 Post-P ninian understanding of the terminology of P nini

K ty yana maintains that there are two types of terms in P nini; krtrima and akrtrima. There are terms in P nini that are completely arbitrary and one may fail to speculate what it implies unless formally informed. Such terms are nadi, ghU[2], TI, etc. According to K ty yana these are said to be artificial terms. Whereas the terms like sarvan ma, avyaya, avas na, etc. are mnemonic or suggestive such that, one who has a lexical repertoire of Sanskrit that would be expected to be able to speculate though not with complete precision, for what does a term roughly stand. For this reason the terms are said to be akrtrima.

Here we can see an attempt to create a meta-terminology pertaining to the terms in P nini. P nini has his own meta-language to describe the facts of Sanskrit. K ty yana by classifying the terms as artificial and non-artificial has initiated a meta-meta-language, to understand the descriptive technique of P nini.

We are going to argue that the classification of artificial and non-artificial does not hold good. It may be insightful but lacks the formal understanding of the semiotics of the technical terms and terminology in P nini or any other system. The A   dhy y of P nini being a formal grammar, the terms in it are bound to be technical and arbitrary and no criterion as semantically suggestive of the function, etc. can come through to say that these are non-artificial. We have to accept the arbitrary and formal nature of the terms anyway. So, this kind of classification is formally not very convincing, though for a traditional pedagogical purpose it may be helpful.

P nini need not be thought of as the only revivalist, even the post-P ninian grammarians like Kondabhatta of the P ninian tradition seem to have developed distinct concepts. And therefore we may further study skeuomorphism. Those however may be left out of the scope of this paper.

3.3 Pierce's Pragmaticism and the technical terms in the P ninian tradition

The meaning of a given concept equals the conceivable consequences of the concept; this is one of the main points in C. S. Peirce's doctrine of Pragmaticism. To put it in the words of Thellefsen (2003: 2)

> *The consequences of a concept are materialized in the related terms of a concept. If we are able to identify all the related terms of a given concept, ideally speaking, we will have total knowledge of the concept. This could be an enormous undertaking since it is impossible in the real world to gain total knowledge of a concept due to the future consequences of the concept. However, the task is made considerably easier since the meaning of a concept is relative to the knowledge domain from which it originates. However, this does not mean that the meaning of a concept is fixed.*

This doctrine of pragmaticism can be an insightful philosophical perspective to understand and note the terminological developments in any knowledge system. So also in P nini we can see the terms as materialized concepts; and in the post-P ninian traditional grammarians' works we can see the consequences of the concepts in several ways. Pragmaticism thus comes through to understand the fundamental concepts in P nini, how these concepts are materialized in the related networks of terms and how these are developed further in the tradition. All this becomes very much relevant to the intellectual history where one may be interested, to put it in the words of Pollock (2008), "to study the discursive deployment of the terms".

This needs a fuller attention and a dictionary of the technical terms in P nini may be thought to be developed which traces the development of the terms throughout the tradition. In this paper, we

---

[2]   Such capitalization is to indicate that the letter is a marker or anubandha in P   ini.

shall briefly see how Skeuomorphism is a device to understand certain terms in the Pāṇinian tradition. Some discussion is already done under *why Skeuomorphism.*

4 Skeuomorphs

4.1 Puruṣa

P.1.4.101. *tiṅas trīṇi trīṇi prathama-madhyama-uttamāḥ*

'each set of three among (the finite verb endings collectively called) *tiṅ* (is called) *prathama* "third (person)", *madhyama* "second (person)", *uttama* "first (person)" respectively.'

This is a *saṃjñā-sūtra* that defined *prathama*, *madyama* and *uttama* (*puruṣa*). Note that there is no provision of defining *puruṣa* and it is we who have been putting *puruṣa* after it.

The terms *prathama* 'first', *madhyama* 'second' and *uttama* 'last' have been borrowed from pre-Pāṇinian authors. Pāṇini does not use the term *puruṣa* 'person' as a technical term, but it is known to Kātyāyana (Vt. VIII on P. 1.4.1). To quote Bhate (1970:47, 48):

> *Uttama: the concept of grammatical person is as old as the Nirukta. There, the terms denoting person are prathamapuruṣa, madhyamapuruṣa and uttamapuruṣa. It appears from the statement made by Yāska that this threefold division of persons was in connection with verbal forms.*
>
> *This concept, which was already established, is accepted by Pāṇini. But, as usual, he reduces this category also to the endings, which are mainly responsible for the division. Pāṇini did not require the term puruṣa besides the term uttama, madhyama and prathama. The personal endings are referred to as tiṅ.*
>
> *Pāṇini defines the term uttama as the third triplet of the tiṅ suffixes. This definition is purely structural. Since the meaning of the term is well known, Pāṇini does not give a semantic definition of the term. This triplet of the first person endings is referred to as uttama, the triplet of the second person endings is referred to as madhyama and the triplet of the third person endings is called prathama.*

To further quote Bhate (1970: 99):

> *Prathama: Pāṇini has restricted the application of this older concept of person to the finite verb endings. Without reference to the term puruṣa he defines the term prathama as the first triplet of the tiṅ suffixes.*

What is notable is that Pāṇini uses the older notion or concept of person but does not use the term *puruṣa,* technically. This very idea of reducing significant grammatical concepts to the (abstract) morphemic units is very typical of Pāṇini.

4.2 Vibhakti

The same mutatis mutandis, is true for the term *vibhakti.*

To quote Bhate (1970: 110-112):

> *"The term means 'division'. In grammar it means 'a case'. B. Leibich has shown that the very first occurrence of the term in the Kauṣitaki-Brāhmaṇa forms the starting point for the development of its technical meaning. The term vibhakti appears in the Kauṣitaki-Brāhmaṇa in connection with the punar ādhāna or replacing of the three sacrificial fires. The six different names where the name of Agni occurs in different cases are described as containing different vibhaktis. Thus the technical meaning of the term is already established here, although the list of all vibhaktis is not yet complete and the order of these vibhaktis is different from the one that we have now.*
>
> *In the Nirukta the seven different cases appear in their fixed order for the first time. Here the term vibhakti makes its first appearance as a technical term. Pāṇini has accepted this*

older concept of vibhakti. But in order to extend the scope of its application he defines it. **The definition points that the term does not refer to any abstract concept like case but to case-terminations and verbal endings also.** *(Emphasis added). Both the case-termination and verbal endings share some common grammatical operations, which P nini, prescribes with reference to the term vibhakti. Similarly a certain group of secondary suffixes is also designated as vibhakti in the A dhy y for the same reason. The term, in this way appears in a wider sense in the A dhy y . It must be noted in this context that although P nini has done away with the concept of case in the definition of vibhakti and reduced it to morphemic units, he still appears to retain the older concept consciously or unconsciously, in terms like pratham , dvit y , etc."*

P.1.4.104. *vibhakti ca*
"also (each set of three is called) *vibhakti*"
As a technical term it has been borrowed from pre-P ninian authors. In the A dhy y the term *vibhakti* is used in three senses:
(1) Case endings and finite verb endings
(2) Case endings exclusively, e.g. P.2.1.6; 8.4.11
(3) The suffixes introduced in the *taddhita*-section from P.5.3.1-27

The s tras pertaining *vibhakti* are:
P.1.4.101. *ti as tr i tr i prathama-madhyama-uttam h*
P.1.4.102 *tany ekavacana-dvivacana-bahuvacan ny eka a*
P.1.4.103 *sUPa*
P.1.4.104 *vibhakti ca*

One thing is clear that triplet-formation, which involves *vacana* (number) and the *puru a* (person) or *vibhakti* (case), is first constructed, though initially for the sake of *ti* is then used for the *sUP* also. And it is notable that *vibhakti* is not simply termed for the *sUP* but also for the *ti* . This, in fact, sheds light on the agreement feature in typology and captures a greater linguistic fact in a shorter convention. In other words, it simplifies in a way such that it as if says that look all is the same for *sUP* and *ti* except that *sUP* are nominal suffixes and *ti* are verbal suffixes, as only they can have the person. One thing is necessary to be aware of at this point of discussion is that there is no idea of distinction between inflection and derivation in P nini. Though here we are not discussing about derivation it is pertinent as far as the explanatory adequacy is concerned which is exhibited by merging the idea of verbal conjugation and nominal inflection into vibhakti in genereal. To quote Lele and Singh (1984):

> *"In Singh and Ford (1980), it was argued that the Graeco-Roman distinction between inflection and derivation is an artifact of the notion of paradigm, entities set up to facilitate the learning of Greek and Latin. The distinction is a pedagogical construct and cannot be justified in a grammar that seeks explanatory adequacy. It is interesting to note that the distinction does not exist chez P nini and that he has in fact been criticized by Max Muller, for example."*

These are the basic s tras for the verbal and nominal word-formation as these actually construct an unfilled chart or a table or rather an abstract generative paradigm, the key of which will be provided as per the form that is to be reached at.

One may point out that what has been mentioned *abstract* in the initial discussion and in the above paragraph, actually, contradicts what is quoted from (Bhate 1970: 110-112), in paragraphs above, where the letters are deliberately highlighted bold as emphasis, to carry out the discussion here. Bhate (1970) holds an idea of abstraction based on binary distinction between abstract against the morphemic. Our understanding is that the morphemic in P nini is further abstract, at another level. We do not reject, nor are we heading towards a refutation of Bhate (1970), as it is a fact that P nini does away with the abstract idea of case and reduces *vibhakti*, as a technical term of morphemic unit. What we say is that these morphemic units are further *abstract*. Our claim or understanding is more in accordance with Subrahmanyam P.S (2007:320) which says:

> *"The … discussion enables us to provide answers for the following two crucial questions: Why did P nini set up* **abstract** *[Emphases added] suffixes to correspond to the tense/mood meanings? Is the substitution of the tense/mood suffix by a personal suffix justified? P nini had to set up the abstract lak ras since there is no particular suffix associated with any particular lak ra. The substitution of a tense/mood suffix by a personal suffix is justified since the change in the form of the personal suffix in association with other features in the form concerned signals the particular lak ra in some cases. Further, P nini had to resort to the technique of substituting a personal suffix to lak ra since, apart from the root; it is the only other morph available in all the lak ras."*

As Subrahmanyam P.S (2007:320) says *lak ra* to be abstract, in the same way mutatis mutandis, we understand the *sUP* and *ti* to be abstract. This actually invites us to a separate research paper as *the abstract in P nini* to detail over the issue. However we close the discussion about the abstract in P nini here and continue with the term *vibhakti*.

Vibhakt is thus all the case-terminations and verbal endings. P nini has defined only *pratham* in P.2.3.46, however the rest of the terms are not defined. By defined, in case of *pratham*, it is only intended that a more detailed, descriptive and meaning-conditioned definition of *pratham* is given. It should be noted that it is not defined by enumeration. It is simply elaborated more than the other *vibhakti*-terms *dvit y , t t y ,* etc. So to sum up, the term *puru a* is not used technically by P nini, but *prathama, madhyama, uttama* are defined, whereas *vibhakti*, unlike *puru a* is used technically, which also includes *puru a* (*ti* -suffixes)in its designation and is defined. But *pratham , dvit y* etc. are not defined. This might be so because it is *pratham* that has an extra-linguistic function of mentioning a word, no other *vibhakti* has such power.

A reason behind not making *puru a* a technical term could be that, it is already labeled *vibhakti* by a following s tra. So needs no more nomenclature. Apart from being redundant it would have marred P nini's favorite *l ghava* or brevity.

Thus the term *vibhakti* becomes non-prototypical in P nini. It thus qualifies to be a skeuomorph as it is only the form of the term that is retained. It has undergone several formal ramifications and these are so much so that we can easily say that the term *vibhakti* in P nini is a skeuomorph of the same older term. It is no longer in P nini a mere division or an abstract case. To sum up the discussion above *vibhakti* in P nini is a bunch of suffixes.

4.3 Sarvan ma
4.3.1 On the term sarvan ma

This is a pre-P ninian term as noted in Bhate (1970: 119). The term in its etymological sense means 'a noun for all'. This etymologically significant term has been used by the grammarians to denote the technical meaning 'pronoun'. Note that there is a point of departure from its lexical

meaningfulness (or what we have called 'etymological significance') to its technical employment in P nini. Though P nini has moved far away from the general meaning of the term, he has retained it. He has not given some technical artificial name to this class.

The term sarvan ma occurs in Nirukta 1.7 as *tva iti vinigrah rth yam sarvan m nud ttam* 'tva a pronoun in the sense of opposition is unaccented.' It should be noted that pronouns are accented by default. It is in some specific sense or some optional forms that they are unaccented. A noun was believed to have two features: a) being accented and b) being declined. Nirukta names tva to be an ardhan man (a half-noun) as it is unaccented and still declined. That does not entail that the nouns that fulfill both these conditions are sarvan ma because then any noun will be a sarvan man. The term sarvan man, therefore, does not mean sarva n ma 'a full name', but sarvasya n ma 'a name for all'.

4.3.2 P nini on the term sarvan ma

He accepts it but does not accept a semantic definition of the term sarvan man, because this would lead to over-application and include the adjectives like *sakala*. He gives a list of words that he labels sarvan ma and thus defines by enumeration. He defines sarvan man as:

P. 1.1.27 *sarv d ni sarvan m ni*
"sarva etc[3]. are called sarvan ma".

The s tras P. 1.1.27 to 1.1.36 deal with pronouns. In the s tras P. 1.1.28 to 36, several other restrictions and facts like optionality etc. are given. We won't look into the each s tra as it is already done in Joshi and Roodbergen (1991:34-44). We shall note some major contrasts in the modern approaches as against P ninian framework towards the understanding of pronouns:

- The P ninian framework does not model or recognize formally, as a part of the theory, pronouns as something anaphoric, but mere forms of inflection (of subantas) that a derivational process (which is described within the framework) licenses.

- Though doing so leads to miss the anaphoric aspects of pronouns, it adds on the front of the economy of classification, such that, all pronouns and the whole of the pronominal inflectional forms are taken care of by the technical designation sarvan man or sarv di, which form a special case of the subantas and thus need no special formal class (other than subantas) to be conceptualized.

- One may claim on this evidence that, as far as the pronouns are concerned, the P ninian treatment to the pronouns is more abda-centric[4].

- It is clear that, atleast here; the inflected form has precedence over anything. In fact, it has been chosen as the defining factor for pronouns in Sanskrit. So for this grammar, pronouns are essentially the specially-inflected nominal (i.e. subanta) items.

So there is a clear-cut purposeful alteration that is consistently exhibited with the significant terms in the grammar. What is noteworthy is P nini altered the terms in such a way that it suits his

---

[3]   The 31 members of the *sarv diga a* are: *sarva* 'all', *vi va* 'all', *ubha* 'two', *ubhaya* 'two', stems ending in -( )*atara* and –( )*atama* (These suffixes are assigned after the stems *kim-, yat-, tad-*, and optionally after *eka-*), *tvat-, tva-* 'other', *nema* 'half', *sama* only in sense of 'all' but not in other senses like 'year' or 'even (ground)', *sima*, 'all, whole', *p rva* 'east, prior', *para* 'subsequent', *avara* 'western, posterior, inferior', *dak i a* 'south, right', *uttara* 'north, higher', *apara* 'other, inferior', *adhara* 'western, inferior', *sva* in the sense of 'own' but not when it means 'blood relation' or 'wealth', *antara* in the senses 'situated outside the town' and 'cloth for wearing (i.e. lower garment)', not for covering the upper part of the body, *tyad* 'that', *tad* 'that', *yad* 'which' (relative pronoun), *etad* 'this', *idam* 'this', *adas* 'that', *eka* 'one' *dvi* 'two', *yu mad* '2nd person', *asmad* '1st person', *bhavat(u)* 'you (honorific), *kim* 'who, what'

[4]   We are deliberately keeping away from saying something like the anaphoric approach to pronouns is more syntactic than this one, as it is a vexed issue as what are precisely the moorings of the Sanskrit grammar : v kya-based or abda-based. Prof. S. D. Joshi and his students argue it to be v kya-based.

descriptive technique that is far more rigorous, formal and adequate compared to the pre-P ninian grammars. In all the instances that we have seen so far, it is clear that P nini is doing away with the older notions and adding substantially new and formal ideas. In most of the cases we see almost a new conceptual stand-point taken by P nini. The alteration or difference is so much so that it is only the form of the terms that has been retained in P nini, the meaning, the definition, the conceptual framework and the descriptive mechanism are considerably new.

We discuss the differences in so much detail not only because it is only then evident how the term in P nini is a skeuomorph of the older, but also because it gives a clear picture of how the terms have been redefined to match the descriptive mechanism. Skeuomorphism emphasizes the fact that the form of the term has been retained. This is noteworthy, as there are also technical terms in P nini that are arbitrary and not attested in the older grammars.

So the only reason we can see why P nini retains some terms is: their use in the older grammars. So to put in formal words, though the conceptual framework was rigorously developed by P nini, the related terms that materialize the conceptual idea were retained as the older ones, though then they did not have any or much functional or semantic significance.

References

Bhate, S.V. 1970. *Pre-P ninian grammatical elements in the P nini's A dhy y* . Unpublished Ph.D thesis, University of Pune, Pune.

Dasgupta Probal 2008. "Knowledge and language in classical Indian linguistics: some observations". Ed. Suresh Raval, G M Mehta, Sitanshu Yashaschandra *Forms of knowledge in India: critical revaluations*. 89-104. Pencraft, Delhi.

Joshi, S. D. and Roodbergen, J.A.F. 1991-2007. *The A dhy y of P nini* (thirteen volumes). Sahitya Akademi, New Delhi.

Sheldon, P. 2008. "Indian Knowledge and the Problem of Early Modernity". Ed. Suresh Raval, G M Mehta, Sitanshu Yashaschandra *Forms of knowledge in India: critical revaluations*. 89-104. Pencraft, Delhi.

Singh, R. and Ford, A. 1980. "Flexion, Derivation et P nini". In K. Koerner, ed., *Studies in the History of Linguistics.* 324-332. Benjamins, Amsterdam.

Subrahmanyam P.S. 2007. "Telugu Grammar in P ninian framework- Some Issues". *Indian Linguistics* Vol. 68 Nos. 3-4 (July-Dec): 315-324. Pune.

Thellefsen, T. 2009. "Semiotics of Terminology: A Semiotic Knowledge Profile." *SEED 3-2* 2003. Retrieved 13 January 2009 from < http://www.library.utoronto.ca/see/SEED/Vol3-2/Thellefsen_3-2.htm>

Appendices
(Appendices are taken from Bhate (1970: 26-271). The places left blank are due to missing pages in the source. The missing pages are: 211, 214, 215, and 216)

Appendix A
Technical terms pre-P ninian in form and meaning and defined by P nini

1. *anud tta*	3. *ap kta*	5. *abhy sa*	7. *avyaya*
2. *anun sika*	4. *abhyasta*	6. *avas na*	8. *avyay bh va*

9. *tmanepada*
10. *mantrita*
11. *rdhadh tuka*
12. *uttama*
13. *ud tta*
14. *upadh*
15. *upasarga*
16. *upasarjana*
17. *ekavacana*
18. *k t*
19. *gu a*

20. *guru*
21. *tatpuru a*
22. *taddhita*
23. *d rgha*
24. *dvandva*
25. *dvivacana*
26. *dh tu*
27. *nip ta*
28. *pada*
29. *parasmaipada*
30. *prag hya*

31. *pratyaya*
32. *prathama*
33. *pr tipadika*
34. *pluta*
35. *bahuvacana*
36. *bahuvr hi*
37. *madhyama*
38. *laghu*
39. *lopa*
40. *vibhakti*
41. *vibh*
42. *sam sa*

43. *sarvan man*
44. *savar a*
45. *sa yoga*
46. *sa hit*
47. *s rvadh tuka*
48. *svarita*
49. *hrasva*

Appendix B
Technical terms pre-P ninian in form and meaning and undefined by P nini

1. *adar ana*
2. *advyupasarga*
3. *anantara*
4. *anatyantagati*
5. *anarthaka*
6. *an r a*
7. *anukara a*
8. *anuv da*
9. *anusv ra*
10. *antartama*
11. *antahp dam*
12. *anv de a*
13. *abhipr ya*
14. *artha*
15. *avagraha*
16. *uttara*
17. *uttarapada*
18. *udaya*
19. *upade a*
20. *upam*
21. *upas a*

22. *upasthita*
23. *upottama*
24. *ubhayath*
25. *c*
26. *o hya*
27. *kriy*
28. *caturth*
29. *chandas*
30. *dvit y*
31. *napumsaka*
32. *nigama*
33. *nirdi a*
34. *pañcam*
35. *p da*
36. *pums*
37. *p rva*
38. *p rvapada*
39. *prak ti*
40. *prak ty*
41. *pratham*
42. *prayatna*

43. *prayoga*
44. *bahulam*
45. *bhavi yat*
46. *bh va*
47. *bh*
48. *br hma a*
49. *mantra*
50. *m rdhanya*
51. *yajus*
52. *yoga*
53. *lak a a*
54. *linga*
55. *vacana*
56. *var a*
57. *v kya*
58. *vidhi*
59. *vibh ita*
60. *visarjan ya*
61. *vaiy kara a*
62. *vyav ya*
63. *abda*

64. *abdakarman*
65. *e a*
66. *atva*
67. *a h*
68. *sa jñ*
69. *sannikar a*
70. *sattva*
71. *saptam*
72. *samartha*
73. *sam napada*
74. *sar pa*
75. *s man*
76. *s m nya*
77. *siddha*
78. *s tra*
79. *str*
80. *sth na*
81. *svara*

Appendix C
Technical terms pre-P ninian in form and undefined by P nini

1. *ap ya*

2. *de a*

Appendix D
Technical terms pre-P ninian in concept

(Technical terms in the present section do not occur in pre-Pāninian works. But the notion they stand for are known before Pānini).

1. *aK*	9. *eC*	17. *cU*	25. *ya*
2. *aC*	10. *eka ruti*	18. *jaS*	26. *lA*
3. *a i*	11. *k p(upadhm n ya)*	19. *jhaY*	27. ---
4. *aL*	12. *kU*	20. *jha*	28. *lI*
5. *sya*	13. *khaY*	21. *U*	29. *lU*
6. *iC*	14. *khaR*	22. *tU*	30. *lU*
7. *it*	15. *K it*	23. *pU*	31. *l*
8. *uccaistar m*	16. *gotra*	24. *pr pti*	

Appendix E
Technical terms defined by Pānini and not found in pre-Pāninian works

1. ---	17. *karmapravacan ya*	33. *dvigu*	50. *lE*
2. *a*	18. *k tya*	34. *nad*	51. *lO*
3. *adhikara a*	19. *gati*	35. *ni h*	52. *vaL*
4. *ap d na*	20. *gha*	36. *ba*	53. *va*
5. *aM*	21. *gh*	37. *bha*	54. *v ddha*
6. *a*	22. *gh*	38. *bha*	55. *v ddhi*
7. *mredita*	23. *aM*	39. *maY*	56. *lu*
8. *iK*	24. *caR*	40. *yañ*	57. *a*
9. *i*	25. *chaV*	41. *yaM*	58. *a*
10. *uK*	26. *jhaR*	42. *yaY*	59. *samkhy*
11. *upapada*	27. *jhaL*	43. *yaR*	60. *sampras ra a*
12. *e*	28. *jha*	44. *yuvan*	61. *sambuddhi*
13. *aiC*	29. *i*	45. *raL*	62. *sarvan masth n*
14. *kart*	30. *ta*	46. *lI* -*a*	63. *sUP*
15. *karman*	31. *tadr ja*	47. *luK*	64. *sU*
16. *karmadh raya*	32. *ti*	48. *luP*	65. *hetu*
		49. *l*	

Appendix F
Technical terms undefined by Pānini and not attested in pre-Pāninian works

1. *akathita*	11. *ant divat*	18. *asarvavibhakti*	27. *kart abhipr ya*
2. *akarmaka*	12. *anyatarsy m*	19. *asiddha*	28. *karmavat*
3. *ati yana*	13. *anyapad rtha*	20. *k nk*	29. *k rya*
4. *adhik ra*	14. *apradh na*	21. *khy*	30. *kriy tipatti*
5. *anabhihita*	15. *ardhahrasva*	22. *dyantavat*	31. *kriy samabhih ra*
6. *an k nk*	16. *aluK*	23. *dh ra*	32. *gu avacana*
7. *anityasam sa*	17. *avidyam navat*	24. *ekavat*	33. *tatk la*
8. *an psita*		25. *ekavibhakti*	34. *tulyayoga*
9. *anude a*		26. *eka e a*	
10. *anuprayoga*			

35. *t t y*
36. *t t y sam sa*
37. *diksam sa*
38. *devat dvandva*
39. *dvirvacana*
40. *nak atradvandva*
41. *naÑsam sa*

42. *nityasam sa*
43. *parar pa*
44. *par gavat*
45. *pumvat*
46. *p rvanip ta*
47. *prak ra*
48. *bh itapumska*

49. *laghuprayatnatara*
50. *vapratyaya*
51. *vartam na*
52. *viprati edha*
53. *vi e a a*
54. *vi e ya*
55. *v ps*
56. *vyavahita*

57. *sakarmaka*
58. *sannatara*
59. *sam n dhikara a*
60. *sam s nta*
61. *sambodhana*
62. *sth nin*
63. *sth nivat*
64. *sthaneyoga*

'GITANJALI': A Study in Lyrical Patterns (Syntax, Diction & Rhythm)

Kusum

Department of Hindi, SCD Govt. College, Ludhiana, India

kusumgopal@yahoo.co.in

Abstract

Rabindranath Tagore said that each language possesses its own rich resources and energies. Although originally written in Bengali, the English translations of 'Gitanjali' are read like original poems. The language used by Tagore as translator is highly emotive, highly charged with the energies of the heart. The French symbolist poet Mallarme said: "*Poetry is written with words, not ideas*". The words used by the poet are remarkable for their sound, sense and suggestion. He was quite aware of the importance of the Liquid Consonants – /l/m/n and/s and/j/. Since the songs in Gitanjali are devotional and divinely inspired, patriotic and have nationalistic strains, the tone, rhythm, pitch, cadence have an inner rhyming. The poems don't have any regular proper meter. They are free but still they have a proper discipline. The present paper proposes to establish how in the early 20th Century Gurudev Tagore proved that translation of poetry is the second most important thing literature. The paper will also explore the linguistic structures used for poetic purposes and creative function. Poetry for Tagore was an act of faith. Translation, transcends, the barriers of language, cultures, continents & communities.

1 Introduction

Tagore, in characteristic imagery, expresses his love for someone whose identity remains a mystery throughout his poetry. In songs he assumes the role of the beloved of God, poetry reaches its highest lyrical intensity. This intensity reveals itself in the melody and music of the language of the songs. We may keep in mind that Gitanjali has to be studied against the background of the devotional movement that flowered in Bengal. The poetic power and the beauty of its imagery can be seen in the works of Chaitanya, Surdas, Jayadeva and Tagore. The rhythm of Tagore's song offerings is the rhythm of renewal. To avoid stagnation in life and the world, God fills it again and again with fresh vigour. Tagore finds God the superb master of music. The poet is astonished at the richness of God's melody. He sees harmony and perfect balance everywhere:

'Ah, thou hast made my heart captive in the endless meshes of thy music.'

2 Discussion

The tone of Tagore's songs reminds one of the poems of Emily Dickinson which are meditative lyrics.

> "**OBSTINATE** are the trammeles, but my
> heart aches when I try to break them.
> Freedom is all I want,but to hope
> For it I feel ashamed.
> I am certain that priceless wealth is
> In thee, and that thou art my best
> Friend, but I have not the heart to
> sweep away the tinsel that fill my room."(Song 28)

It has a deeply personal tone and seems to bring out the conflict that has shaped Tagore's life.

Narada Says: "But that state of supreme love and immortality is made possible only by giving up the attachment to objective reality of the world"

<div align="right">(Bhakti Sutras, III. 35)</div>

Pride, vanity and other vices are to be given up आभिमान दाास्मक त्याज्य

<div align="right">(Narada Bhakti Sutras, IV. 641)</div>

The Upanishads compare the spiritual path to the razor's edge: 'Sharp as is the edg e of razor lies the path, say the wise.'

Tagore finds that safest path is the path of love and devotion (Bhakti sutras).

Radhakrishnan Says: "Tagore beautifully depicts how an enthusiastic surrender to the spontaneity of natural scenery leads a man to his goal…"

The songs are an attempt to capture the COSMIC rhythm of life-Time, Change and Death. Then the renewal: 'Readiness is all'.

<div align="center">

In song,
I KNOW that the day will come when
my sight of this earth shall be lost, and
life will take its leave in silence, drawing
the last curtain over my eyes.

</div>

<div align="right">(SONG-92)</div>

The poet speaks of after-life:

<div align="center">Yet stars will watch at night, and morning rise as before……</div>

For Gurudev Rabindranath Tagore- "Poetry is an act of faith. So was translation. As poet and translator, he believed that each language possesses its own creative resources and energies. This is the reason why the songs and poems in Gitanjali read like original creations. The language used by the poet is the language of emotion, rhythm, lyrical intensity charged with instinctive movement.

<div align="center">

"**LIFE** of my life, I shall ever try to
keep my body pure, knowing that thy
living touch is upon all my limbs
I shall ever try to keep all untruths
out from my thoughts, knowing that
thou art that truth which has kindled
the light of reason in my mind.
I shall ever try to drive all evils away
from my heart and keep love in
flower, knowing that thou hast thy seat
in the inmost shrine of my heart
and it shall be my endeavour to
reveal thee in my actions, knowing it
is thy power gives me strength to act".

</div>

<div align="right">(SONG – 4)</div>

The French symbolist poet Mallarme said, "Poetry is written with words, not ideas." The words used by Tagore are remarkable for their sound, sense and suggestions. He is deeply aware of the importance of liquid consonants /L/M/N:

e.g:-

<div align="center">

MY Desires are many and my cry is
Pitiful, but ever didst thou save me by

</div>

hard refusals; and this strong mercy
has been wrought into my life through
and through.
Day by day thou art making me
Worthly of the simple, great gifts that
Thou gavest to me unmasked --- this sky
And the light, this body and the life
And the mind ----- saving me from perils
of overmuch desire.
There are times when I languidly
Linger and times when I awaken and
Hurry in search of my goal; but cruelly
Thou hidest thyself from before me.
Day by day thou art making me
Worthly of thy full acceptance by refu-
sing me ever and anon, saving me from
perils of weak, uncertain desire.

(SONG – 14)

e.g. -

I HAVE had my invitation to this world's
festival, and thus my life has been
blessed. My eyes have seen and my
ears have heared.
It was my part at this feast to play
upon my instrument, and I have done
all I could.
Now, I ask, has the time come at
Last when I may go in and see thy face
And offer they my silent salutation?

(SONG-16)

These lines from the songs are highly lyrical. The poetic language, its diction, rhythm, pitch and tone is to be marked:

e.g. - **LIGHT**, my light, the world filling light,
the eye kissing light, heart sweetening
light!
Ah, the light dances, my darling, at
the centre of my life ; the light strikes,
my darling, the chords of my love; the
sky opens, the wind runs wild, laughter
passes over the earth.
The butterflies spread their sails on
the sea of light. Lilies and jasmines
surge up on the crest of the waves of
light.
The light is shattered into a gold on
every cloud, my darling, and it scatters
gems in profusion
mirth spreads from leaf to leaf, my
darling, and gladness without measure,
the heaven's river has drowned its
banks and the flood of joy is abroad.

(SONG – 57)

e.g. -

> **PLUCK** this little flower and take it,
> delay not! I fear lest it droop an
> Drop into the dust.
> It may not find a place in thy gar-l
> and, but honour it with a touch of
> pain from thy hand and pluck it. I
> fear lest the day end before I am
> aware, and the time of offering go by.
> Though its colour be not deep and
> Its smell be faint, use this flower in
> Thy service and pluck it while there
> is time

(SONG – 6)

e.g. -

> "Lifeof my life, I shall ever try to
> keep my body pure, knowing that thy
> living touch is upon all my limbs
> I shall ever try to keep all untruths
> out from my thoughts, knowing that
> thou art that truth which has kindled
> the light of reason in my mind.
> I shall ever try to drive all evils away
> from my heart and keep love in
> flower, knowing that thou hast thy seat
> in the inmost shrine of my heart
> and it shall be my endeavour to
> reveal thee in my actions, knowing it
> is thy power gives me strength to act".

(SONG - 4)

In song,

> **IS** it beyond thee to be glad with the
> gladness of this rhythm? to be tossed
> and lost and broken in the whirl of this
> Fearful joy?

The poet describes in a rhythmic rush the quick passage of time Seasons change, all things rush on, no power can hold. Life goes on:

All things rush on, they stop not, they look not behind, no power can hold them back, they rush on.

These songs have an organic pattern and design. What binds them together is the quest motif. In these poems nothing happens outside. Mind is the theatre where action takes place. The language of the poems has its roots in Indian culture, particularly Hindu way of life. Tagore's search for immortality and identity makes these poems timeless and in ultimate analysis language becomes a gesture.

He employs the same stream of life that determines vision of the divinely inspired poems, their devotional, patriotic and nationalistic and their tone and temper. The poet has used inner rhyming as

a poetic device. The language is simple, conversational, effortless, natural, familiar and of day to day nature. Take these lines:

e.g. -

> LEAVE this chanting singing and
> telling of beads! Whom dost thou
> worship in this lonely dark corner of a
> temple with doors all shut? Open
> thine eyes and see thy God is not before
> thee!
> He is there where the tiller is tilling
> The hard ground and where the path-
> Maker is breaking stones. He is with
> them in sun and in shower, and his
> garment is covered with dust. Put off
> the holy mantle and even like him come
> down on the dusty soil!
>
> (SONG – 11)

e.g. -

> IF thou speakest not I will fill my
> heart with thy silence and endure it. I
> will keep still and wait like the night
> with starry vigil and its head bent low
> with patience.
> The morning will surely come, the
> darkness will vanish, and thy voice pour
> down in golden streams breaking through
> the sky.
> Then thy words will take wing in
> Songs every one of my birds'
> neste, and thy melodies will break forth
> in flowers in all my forest groves.
>
> (SONG – 19)

This whole poem deserves mention in full:-

> WHERE the mind is without fear and
> The head is held high;
> Where knowledge is free;
> Where the world has not been broken
> Up into fragments by narrow domestic
> Walls;
> Where words come out form the
> Depth of truth;
> Where tireless striving stretches its
> Arms towards perfection:
>
> (SONG – 35)

Most of the songs are prayers. The feelings and thoughts are universally shared. These prose poems form a part of universal human heritage. In a unique poetic language the poet has created a

new idea and created a form within form. These songs have the openness and vastness of Alexander Pushkin's poetry. Consider these lines-

> **ON** the seashore of endless worlds
> children meet. The infinite sky is
> motionless overhead and the restless
> water is boisterous. On the seashore
> of endless worlds the children meet
> with shouts and dances.

(SONG -60)

The linguistic structures that Tagore uses make the poems real works of art. Each song has an organic unity. It is a perfectly organized piece of art.

e.g. –

> **THE** same stream of life that runs
> Through my veins night and day runs
> Through the world and dances in
> Rhythmic measures.

(SONG – 69)

In contemporary critical practice, the reading of Gitanjali becomes an act of rereading as in language nothing in final, certain and stable. No doubt, we live in language. Almost all literary critics and scholars since Aristotle emphasise structures as the crux of literature but the point is that language with its magic and dream like quality will never be conquered. Herein lies the eternal charm of the songs of Gitanjali. Its language is sublime and has decorum.

It may be observed that the language of Gitanjali is loaded with cultural and literary potential.It extends meanings into a larger world. The language has rare balance, tone, syntactic structure, pitch and rhythm. It would be quite appropriate here to mention the stylistic quality of the poems, particularly Rass/ Dhvani / Vakrokti.

Paul Valrey says that "In poetry the form and content are written in wedlock."

According to the Sanskrit Scholar **Bhama** – "AlamKar is the most essential element in poetry."These gunas can be clearly seen in the songs of Gitanjali.

According to **Vaman**-"The stylistic refinements are the soul of composition, the figures of speech are not essential refinements,they are properties of the various styles;"

Tagore has used many styles. His songs are marked by appropriate syntactic structures, proper gender and number, adjectives, prepositions, tense and proper thoughts.

I may say that in my brief presentation, I have pointed out certain areas that can be investigated by a researcher by using tools available in contemporary critical practice. Language of the songs of Gitanjali has many currents that embrace a philosophy that can be summed up in the immortal

words of Narada: सा त्वास्मन परम प्रम

That, verily, is the nature of Supreme Love.

3 Conclusion

To move towards the conclusion, Tagore is simply magnificent. The link he tries to raise with the divine through his impenetrable but at the same time fathomable artistic creation, is fabulous. The internal involuntary willingness to glue the supremely enormous in the pages of his immortal creation, with fastidious stress on language, diction and rhythm, has captivated many a hearts and bowled over many minds.

Reference

Tagore,Ravindranath.1957.*Gitanjali*. London Macmillan.

Chakravarty, Amiya. (Ed.).1961. *The Artist*, *A Tagore Reader* by New York : Macmillan Co. port.

Dutt, Indu. (tr.) 1962. *The Diary of a Western Voyage,* , Asia, Bombay.

Chandhari, Indira Nath.1984 . Comparativism and Translation Studies(Bengali, Hindi, English). Paper presented at the first International Conference on Literature in Translation, 15-20 March,1984. Jawaharlal Nehru Univ., School of Languages, New Delhi.

Das Gupta, Prima Kanti.1982. Problem of Translation between Sanskrit, Bengali and Assamese. Paper Central Institute of Indian Languages,Mysore,13-16 May.

 Despatie, Gerard S.J.1967. Modern Linguistic Research applied to the Process of Translation Columbia University.

Dass , Kavi Chandra. (Tr.) 2000. *NARADA Bhakati Sutras (Mysteries of Divine Love), Original Text and Interpretation* (74) Mumbai: .

Draper,H.1962.The art of Re-Creation: In Defence of Verse Translation. South Atlantic Quarterly.

Murthy,V,Rama. 1972. *Tagore's Gitanjali: A Commentary*. Doaba House, Delhi.

Some Methodological Observations of Sociolinguistic Fieldwork: a Case Study from Maharashtra-Karnataka Border

Arvind Jadhav

Y.C. College of Science, Karad, India (MS)

lecturer.arvind@gmail.com

Abstract

This paper illustrates some methodological considerations to undertake linguistic fieldwork in general and sociolinguistic in particular. This case investigates the morpho-syntactic variation in the multilingual situation at the Maharashtra-Karnataka border area. The paper looks at the consideration that we needed to comply with to form a statistically justifiable sample. It also raises the question of area specific variables for sampling. What we need to look at is how each variable will give a different but better insight for theoretical perspectives and their application in the field.

Introduction

Language convergence process that occurs due to prolonged language contact situation has been a topic of interest in Sociolinguistics. Other than a situation of prolonged language contact, language convergence occurs due to immigration especially in the case of tribal languages. In India, prolonged language contact has been the major cause of language convergence. One of the first reported cases of this type was that of Kupwar in Sangli district (Gumperz and Wilson, 1971)..

Similar sociolinguistic processes are found in many more villages around the border regions in India. At the Mahrashtra-Karnataka border, three languages (Marathi, Hindi-Urdu, and Kannada) have been coexisting for centuries in villages such as Arag, Bedag, Ugar, Abdul Lat, Kagal, Kagwad and Kudachi. No published work is so far found after Gumperz' research and hence there is a need to undertake this study to know the present linguistic change in the given area.

Defining the Boundaries

I) Why Kagwad and Kupwad?

Kagwad is a village in the state Karnataka and at present situated exactly on the Maharashtra-Karnataka border after the reorganization of the states on linguistic basis. It was in Mysore state before that. More than half of Kagwad border area is occupied by Marathi speaking villages of Maharashtra and the rest of by Kannada speaking villages of Karnataka.

Kupwad is a village in Maharashtra state at present. In 1961, the village had only a population of 6,000. It crossed 65,000 in 40 years. Major changes have come about in the village because of industrialization in Kupwad, especially regarding settlements of the various communities. Education has also brought about some more changes.

1) Languages: Three language communities are found in both the villages in the pilot survey. Marathi, Kannada and Hindi-Urdu have coexisted for many centuries. Some Telugu speakers also live in both the villages. They are very less in number and not to be included in this study. The inception of one more language (Telugu) will diminish the focus of the study and result in a situation where we will not able to control the variables. Marathi and Hindi-Urdu belong to the

Indo-Aryan language family, whereas Kannada and Telugu belong to the Dravidian language family.

2) Population: The population of Kagwad is 13,299 with 6,753 males and 6,546 females (Census of India, 2001). The population of Kupwad has been increasing fast as seen from the census reports of India (1961-6798, 1971-9351, 1981-11716, 1991-33313 and 2001-67136). It has many reasons. The Kupwad was a village merged in Sangli-Miraj-Kupwad Municipal Corporation cluster in 1998.

3) Distance from Border: Kagwad is 00 km and Kupwad is 20 km away from Maharasthtra-Karnataka border. Ultimately, 20 km is the distance between both the places i.e. Kagwad and Kupwad.

4) Medium of Instruction: No formal education is available in Kannada medium in Kupwad. The people from Kupwad are not familiar with Kannada script, nowadays. Kagwad has both the mediums i.e. Kannada and Marathi. Being a semi-urban area, in both the places transport and communication facility is good. TV channels, radio stations, internet facility is same in both the places.

Kupwad and Kagwad may have started with similar sociolinguistic situations. Now, as the two villages are in two different states, the language convergence process is very likely to have progressed differently.

II) Which Linguistic Items?

The data will be collected from two villages i.e. Kagwad and Kupwad. While defining the topical scope, the study has been restricted to the morpho-syntactic level only, keeping into mind the vast sample size for the data collection. The following morpho-syntactic elements will be investigated:

1) Gender agreement
2) Number agreement
3) Person agreement
4) Non-finite VPs and Compound verbs
5) Copula construction
6) Use of subordinating conjunction; that (*'KI'*)
7) Interrogative verbal suffixes

During the research, some more morpho-syntactic elements will be included if they were found interesting to this study.

Defining the sampling universe

1) Area for case study: The speakers from Kagwad will be interviewed for case study as per parameters and the speakers from proper Kupwad (gavthan) will be selected for interviews and detailed study.

Areas selected for case study: Kagwad (Karnataka) and Kupwad (Maharashtra).
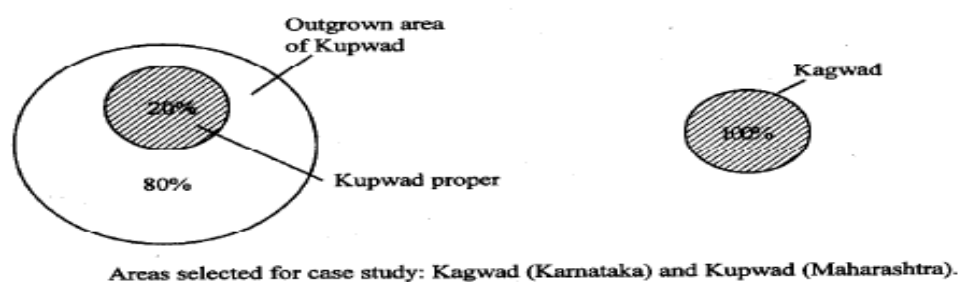
2) Justification: The areas around Kupwad have been growing rapidly from the east, the south and the west. Sangli; the district at the western side and Miraj the *taluk* at the east are growing in the direction of Kupwad. For the selection of speakers, the outgrown areas with one unit called Bamnoli have been kept aside that comprises of 80 per cent of the total population of Kupwad. The rest of the 20 per cent of the population remains in proper Kupwad (approximately, 13427) which is called as gavthan of Kupwad. The families living in proper Kupwad have been there for more than 40 years, and the residents of proper Kupwad are there for two or more generations. The residents of Bamnoli (approx. 2000) are the rehabilitated people because of the Koyana Electricity Project and hence not included in the study. The people residing in the extension or outgrown area belonged to different parts of Maharashtra and Karnataka with different dialects of all the three languages.

Labov rejected persons from his sample as they were not "native speakers" of the dialect or of the language in New York City. Panye's work (1980) suggests that children whose parents come from different dialect may not acquire the structural patterns of the dialect spoken by the community into which they are born.

Selecting the subjects

To select subjects from both the villages to be representative of the districts and the data to make statistically sufficient was the first challenging issue. We had two villages. The data samples are decided to be designed for villages keeping all the sociolinguistic factors in mind. In a sociolinguistic research study, we find the number of speakers is being determined according to the aims and objectives and as situation demands. It has been argued that large samples are not necessary for linguistic surveys as for other surveys. Labov (1966, 180-1) suggested that linguistic usage is more homogeneous than many other phenomena studied by surveys**.** While conducting ethnographic survey of Dharavi, a slum in Bombay, the researcher kept 62 samples for interviews. Out of which 54 were among the slum population and 8 belonged to the resource personnel. This sample was based on a non-random sampling procedure. That was an exploratory survey.

The practical considerations play a role in determining the sample size. The landmark variationist studies were based on samples that were quite small by general socio-scientific standards. Labov's New York study comprised of 88 speakers; and is Trudgil's 60 Norwich speakers (Trudgill, 1974.); 69 speakers selected for sample among 200 in the Detroit high school study (Eckert, 2000);

Susheela Thomas determined 96 speakers in her Kasargod case study on language contact and language convergence (Thomas Susheela, 1995), etc.

In the present study, unlike other studies, **language** has also been considered an important factor to design the samples. As in both the villages Marathi, Kannada and Hindi-Urdu co-existed, the speakers have any one of them as their mother tongue. So, all the languages are given equal importance while selecting the speakers. We took 96 speakers for interview, who were divided into two. It makes 48 for each village. It was also further divided according to **sex** and we got 24 males and 24 females for each the village.

The **age** has also been taken as a major sociological determinant and the sample was divided into two age groups from the two generations. The adolescents (age 16-18) are compared with that of adults of their parent's age (40-55) as suggested by Gordon. Then we have 12 adults and 12 adolescents in each village.

The next variable was **education.**. The 10th and above can be considered the marking line for both the groups i.e. adults and adolescents. The first year junior college students selected for educated adolescents are to be interviewed in both the villages. For the adults, along with education, socio-economic status has been taken into consideration at the same time. The servicemen are selected for interviews who have completed their education at least till tenth. For the uneducated 7th and below determined the demarcation line. The adolescents from the same age group (16-18) having no formal education or below 7th selected for interviews and at last for the uneducated adults along with their socio-economical status in mind, the farmers were selected for the interviews who are having no formal education or below the 7th .

The distribution of total 96 speakers is given in Table 1 and Table 2.

Sr. No.	Factors	Villages	
		Kagwad	Kupwad
1	Age	24 Adolescents 24 Adults	24 Adolescents 24 Adults
2	Sex	24 Males 24 Females	24 Males 24 Females
3	Education	24 Educated 24 Uneducated	24 Educated 24 Uneducated
4	Mother Tongue	16 Marathi 16 Kannada 16 Hindi-Urdu	16 Marathi 16 Kannada 16 Hindi-Urdu

Table 1: Speakers from Kagwad and Kupwad:

Sr. No.	Number of speakers	Description
1	2	Male Adolescents Educated (XI class)
2	2	Female Adolescents Educated (XI class)
3	2	Male Adolescents Uneducated (7th or less)
4	2	Female Adolescents Uneducated (7th or less)
5	2	Male Adult Educated (Service Sector)
6	2	Female Adult Educated (Service Sector)
7	2	Male Adult Uneducated (Farming)
8	2	Female Adult Uneducated (Farming)

Table 2: Speakers of Kannada/Marathi/Hindi-Urdu as mother tongue in Kagwad/ Kupwad:

Method of data collection

The data will be collected through different methods like personal interviews (personal /group interviews) and questionnaires. Participant's observation and investigator's long time involvement is considered as prime importance for an in-depth knowledge of the societal structure and socio-cultural practices in the given communities in the selected area.

Method of data analysis

As the study focuses on the morpho-syntactic elements, the structural grammars will be considered for an analysis of the recorded data. The local varieties of concerned languages will be compared with that of the standard languages.

Some initial steps

After the research design, the investigator spent ten days staying at the boys' hostel in Kagwad. And visits were paid to Kupwad for more than six months intermittently. He tried to involve in the society as possible as he could and recorded some samples of natural conversation to get an idea and make strategies for the required data collection. He recorded data at some public places i.e. at the bus stop, college area, praying places, squares, hotels, shops, on the roads, in lanes, market or bazaar and in homes, hospitals. Some personal interviews were also recorded that can contribute to know some interesting details. For example, in an interview in kagwad, an ex-village-panchayat member gave all the details about a festival and the myth related to it. In Kupwad, one 83 year-old Muslim who was interviewed gave some depiction of Gumperz and him and he claimed that he knew Gumperz and now couldn't recollect all the details. One more retired head master has been recorded and revealed that the parents of him were the host of Gumperz and Wilson. They stayed in house for three months. He also gave some more details about how Gumperz and Wilson would behave with natives, what kind of questions they used to ask, etc. This kind of information could never be extracted from any resources.

While collecting data, some primary observations made by the fieldworker at a very initial stage are as follows. They will help in further study to some extent.

1. Patience is the supreme tool in fieldwork studies.
2. It is a time consuming activity: the field worker should not make any haste.
3. The field worker should be frank and open minded, his being reserved will not be of any use.
4. He should always take all the apparatus of recordings and diaries with him to have chance anytime to record the natural conversation or an important issue. You can capture the speech event any time.
5. He has to check the instruments and make sure it is ready, all the time.
6. Fieldworkers have to fit anywhere and somehow record a conversation. They should be flexible.
7. Fieldworker should update the plan of his work now and then.
8. Always keep in mind not to talk much on sensitive issues. For example, some people do not know much of a particular issue, but you are not there to teach or preach anything.
9. Much people are more demanding and they want to know almost everything about you and your work; without ignoring them try to satisfy within a minimum time instead of making it vague. This will lose your time.
10. Some others are tongue-tied; they speak no word at all.
11. Control of unwanted noises can help to collect clear and audible data.

1 Some theoretical questions

This would be the representative example of a sociolinguistic study at the border areas in India. Some theoretical questions arise regarding the methodology for such studies are as follows:

1. Can the methodologies from bi/multilingual situations elsewhere in the world be applicable in the fieldwork studies to be undertaken in India?
2. How to control different languages, different linguistic and socioeconomic factors at the same time? And if we restrict all the factors to a certain degree, is it research worthy to know the process of convergence or to define variation in language and responsible socio-economical factors?
3. How to know which socio-economical factor is more responsible than the other?
4. How to define the locale for case studies like Kupwad?
5. How far the statistics or mathematical figures give the best results in studies like language variation?
6. What exactly do we mean by statistically sufficient data or justifiable data, as the terms used in the books on methodology in sociolinguistics?
7. Is there any need of standardization in methodologies for initiating sociolinguistic studies in India?

While concluding this paper, I think, to get the answers for all the above and other questions regarding sociolinguistic studies, the urgent need is to undertake more studies considering different sociolinguistic factors separately. Each one of them will give a different but better insight for theoretical perspectives and their application.

References:

Gumperz and Wilson. *'Convergence and Creolization: A case study from Indo-Aryan/Dravidian border in India'.* In Hymes D. (Ed.). Piginization and Creolization of Languages. Cambridge: CUP, 1971.

Gumperz and Wilson. *'Convergence and Creolization: A case study from Indo-Aryan/Dravidian border in India'.* In Hymes D. (Ed.). Piginization and Creolization of Languages. Cambridge: CUP, 1971.

Kupwad is the same of Kupwar as used by Gumperz and cited in further studies of language contact. This spelling has preferred because it is used it is used in the entire written official documents, later one is not in use.

Gumperz and Wilson. *'Convergence and Creolization: A case study from Indo-Aryan/Dravidian border in India'.* In Hymes D. (Ed.). Piginization and Creolization of Languages. Cambridge: CUP, 1971.

Census of India. Director of the Census Operations, Maharashtra. Sangli District: Villages and Town Directory, 1961, 1971, 1981, 1991 and 2001.

Labov, W. The Social Stratification of English in New York City. Washington DC: Center for Applied linguistics, 1966. In Lesley Milroy and Mathew Gordon, 2003. Sociolinguistics: Methods and Interpretation. New York: Blackwell publishing, 2003.

Payne, A. C. Factors controlling the acquisition of Philadelphia dialect by out-ofstate Children. 1980(143-178). In Lesley Milroy and Mathew Gordon, 2003. Sociolinguistics: Methods and Interpretation. New York: Blackwell publishing, 2003.

Labov, W. The Social Stratification of English in New York City. Washington DC: Center for Applied linguistics, 1966 (180-81). In Lesley Milroy and Mathew Gordon, 2003. Sociolinguistics: Methods and Interpretation. New York: Blackwell publishing, 2003.

Rajyashree K. S. An ethnographic survey of Dharavi (a slum in Bombay). Mysore: Central Institute of Indian Languages, Manasagangotri, 1986. **.**

Thomas Susheela. Language Contact and Convergence: A case study in Kasargod. (Ph. D. Thesis). Mysore: University of Mysore, 1995.

Lesley Milroy and Mathew Gordon. Sociolinguistics: Methods and Interpretation, 2003.

Temporality in Bengali: A Syntacto-Semantic Framework

Samir Karmakar
Indian Institute of Technology,
Kanpur, INDIA
samirk@iitk.ac.in

Abstract

The paper integrates the syntax and semantics of temporality in Bengali. Temporality in a language is encoded in the following grammatical categories: verb, tense, aspect, and adjunct. The semantic interpretation of temporality is discussed in terms of situation types, temporal location, continuity and boundary. The paper integrates the syntactic and semantic aspects of temporality to propose a syntacto-semantic interface, with a special reference to the Bengali language data.

1 Introduction

The paper discusses the way an inflected Bengali verb form can be represented to capture the syntactic and semantic features of temporality presented in a sentence. I would first analyze the structure of the Bengali verb morphology in order to arrive at the basic syntactic pattern. Then it would be shown how the different layers of this syntactic pattern correspond to the semantic information relevant to temporality. Finally, the paper would propose a syntacto-semantic framework to represent the temporal information.

2 Syntax and Semantics of Temporality in Bengali Verb Morphology

The syntax of Bengali verb morphology has the following structure:

1. V-aspect-tensei-personi
    For example: gey-ech-il-o
    sing-perf-past-3past
    (s)he had sung (a song).

Tense and personal markers are in agreement in Bengali. The tense information encoded in Bengali personal marker helps the hearer to retrieve the information about the temporal location in a construction like habitual, in a construction that lacks overt tense marking.

2. Retrieving tense information in case of habitual construction:
    khel-t- -o
    play-habitual-past-3past
    (s)he used to sing (classical Hindustani).

In (2), the information of temporal location is retrieved from the tense sensitive personal marker. The lack of overt tense marking is represented by ' '. In this case, the tense sensitive personal marker helps in locating the participants of a situation in time.

To propose a syntacto-semantic framework of temporality in Bengali, one needs to identify and analyze the relevant structural units in greater detail than the traditional grammarians have done. For example, as per the tradition perfective represents a completed situation and is represented by -ech- in Bengali. Instead of following this traditional view, the paper considers the concept of

perfectivity in terms of continuity and boundary. In Bengali a perfective situation represents a bounded continuum on the time line, in contrast to the unbounded continuum in case of an imperfective situation. As a consequence, perfectivty is considered as a complex category. This understanding of perfectivity in Bengali motivates me to decompose the corresponding marker -ech- in -e- and -ch- with the following semantic design features: The appearance of -e- in the structure marks the bounded feature for the situation, and -ch- specifies continuity. In Bengali -ch- represents imperfectivity because of conveying a sense of continuity. Syntactically, -e has been conceptualized as participle. The plausibility of this structural analysis could be noticed in the following example:

3. -e- and -ch- marking boundary and continuity respectively:
   khel-e phir-ch-il-o
   play-part return-impf-past-3past
   Having played (football), (s)he was returning (home).

The gerundive use of -e- in example (3) shows that the inflected verb form (= khel: play) represents the sense of a completed situation: Sense of completion presumes the bounded reading of a situation. On the other hand the appearance of -ch- in (3) ensures the imperfective interpretation that is the unbounded reading.

Therefore, the structure of the inflected Bengali verb represents the following syntacto-semantic correspondence:

4. Syntax-Semantics correspondence of temporality in Bengali verb morphology:

V	aspect		tense$_i$	person$_i$
adicity[1]	perfect/imperfect		past/present/future	$1^{st}/2^{nd}/3^{rd}$
situation type	boundary	continuity	temporal location of the situation	temporal location of the participants

In the following section I would elaborate how the information relevant to *situation type*, *boundary*, *continuity* and *temporal location* can be represented systematically.

3 Organizing Bengali Syntax and Semantics of Temporality

The above mentioned synatcto-semantic description of the Bengali verb morphology is further systematized by structuring the representation in terms of *inner* and *outer* layers: (i) The inner layer of the representation concerns about the syntax and semantics of the verb and the adjuncts. A verb encodes information of the argument structure and the situation types. However it is also noticed that the semantics of an adjunct often modifies the default situation of the verb (Reichenbach 1947: 302).[2] Adjuncts are guided by a set of co-occurrence constraints that allow them to occur with some situation type while barring the others (ref. section 3.3). (ii) The outer layer deals with the syntacto-semantic behaviour of tense and aspect. Since the Bengali personal markers are in agreement with the tense markers, they also contribute in extracting the temporal information from a text.

---

[1]   In the symbolic tradition *adicity* is defined as *n-place function*.
[2]   *Adjuncts* are considered as *higher order function* in symbolic logic.

Representing Syntactic Information Relevant to Temporality

The syntactic information is modeled according to the proposal of the Minimalist Program, where the underlying principle is *X-bar theory* (Chomsky 1995: 172-186)[3], i.e.

[XP Spec(…), [X' Head(X), Comp(…)]]

The symbol 'X' stands for a variable belonging to the class of syntactic categories. A class of syntactic categories consist of *noun* (N), *verb* (V), *adjective* (Adj), *adverb* (Adv), *complementizer* (C), *subject agreement* (AgrS), *tense* (T), etc. In addition, I have introduced *aspect* (Asp) as an abstract syntactic category. If it appears in the inner layer of the frame, it would be considered as *lexical aspect*; otherwise it would be interpreted as *grammatical aspect*. Hence these two variants of 'Asp' are structurally constrained. 'X" stands for the intermediating projection of 'X', and 'XP' is the maximal projection. 'P' stands for 'phrase'.

Representing Semantic Information Relevant to Temporality

Following Reichenbach (1947: 287-301), the information about the temporal location could be captured in terms of relation holding between *speech time* (= S), and *reference time* (= R). On the other hand, the relation between reference time and *event time* (= E) gives an idea about the grammatical aspect, by specifying *boundary* and *continuity*. These two relations are popularly known as *first* and *second referencing,* respectively. Finally to represent the information of the situation type I have adopted the following approach of Dowty (1979: 71-132) to represent Vendler's four situation types (1957).

5. Situation Types:

Types	Definition
State:	$V_n(a_1, a_2,…,a_n)$
Activity:	$DO(a_1, V_n(a_1, a_2,…,a_n))$
Accomplishment:	$DO(a_1, V_n(a_1, a_2,…,a_n))$ CAUSE(BECOME($V_n(a_1, a_2,…,a_n)$)))
Achievement:	BECOME($V_n(a_1, a_2,…,a_n)$))

The information pertinent to the *temporal location*, *boundary* and *continuity* constitutes the *outer layer* of the frame; whereas *situation type* along with the *argument structure* constitutes the inner layer of the frame (Verkuyl 1989, 1993; Arche 2006). *Inner layer* of the frame represents situation type in terms of *duration*, and *telic* features. *Outer layer* represents information relevant to *temporal location*, *boundary* and *continuum* in terms of the *dyadic predicate relations*[4], like *precedence* and *overlap*:

1 the relation between speech time and reference time to represent different temporal locations, like past, present and future; and

2 the relation between reference time and event time to construct either *perfective* or *imperfective* interpretation.

The framework outlined in figure 1 integrates syntactic and semantic conceptualization in a unified representational system in order to deal with the temporality.

---

[3]   Chomsky proposed the *X-bar theory* first in *Lectures on Government and Binding* (1981).
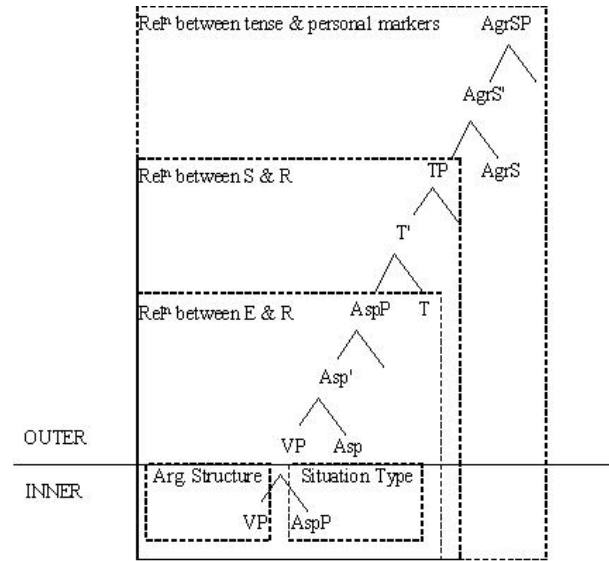[4]   I adopt the term *dyadic predicate relation* from Arche (2006: 148).

Fig 1: Syntacto-Semantic Frame of Temporality in Bengali

In the following section the semantic contribution of adjuncts will be discussed to show how they specify the situation type at the sentential level.

4 VP internal Aspect

Headings: The adverbial adjunct modifies the default situation type of a verb, thus coercing the verb semantics. Such a modification is proposed to be dealt VP internally.

6(a)	panero	miniT-e	gey-ech-il-o
	Fifteen	minute-part	sing-perf-past-
			3_{past}

(s)he had sung (the song) in fifteen minutes

6(b)	*panero	miniT-e	gAi-ch-il-o
	fifteen	minute-part	sing-impf-past-
			3_{past}

*(s)he was singing (the) song in fifteen minutes

The sentence (6a) from Bengali has two readings: (i) the act of singing was accomplished *in fifteen minutes* (*accomplishment*); and, (ii) the act of singing was started *in fifteen minutes* (*inception*). What seems to be interesting in both these two interpretations is a sense of *boundary*: in case of (i) it is *ending*, and in case of (ii) it is *beginning*. In Bengali, *singing* is an activity verb. In conjunction with the perfective marker it projects a sense of *completion*, hence having a boundary. In addition the semantics of *panero miniT*, marked with –*e*, has a sense of boundary as per our discussion. Hence, the expression *panero miniT-e* satisfies the semantic expectation of *gey-ech-il-o*. However in case of (6b) the semantics of the adjunct is anomalous to the semantics of the inflected verb form which in turn results in an ungrammatical construction. The act of singing in (6b) is marked imperfectively, so lacking a bounded interpretation, in contrast to the bounded interpretation of the adjunct. However, the following construction is grammatically well formed:

7(a)	panero	miniT	dhar-e	gAi-ch-il-o
	fifteen	minute	hold-part	sing-perf-past-3$_{past}$

!(s)he was singing (the song) for fifteen minutes.

In contrast to the semantics of *panero miniT-e* of (6b), in example (7a) *panero miniT dhar-e* has a sense of *bounded continuum*. More specifically, the phrase *panero miniT dhar-e* not only conveys the sense of the boundary but also communicates the sense of immediately preceding interval, in contrast to the semantic behaviour of *panero miniT-e* which only communicates the sense of boundary. As a consequence, the semantic expectation of the inflected verb form of (7a) is satisfied. Due to having a sense of bounded continuum, the perfectively inflected verb form in conjunction with the *panero miniT dhar-e* does not result in an ambiguity as was the case with (6a) discussed earlier.

7(b)	panero	miniT	dhar-e	gey-ech-il-o
	fifteen	minute	hold-part	sing-perf-past-3$_{past}$

!(s)he had sung (the song) for fifteen minutes.

The discussion shows how the semantics of adjuncts and verbs interact with each other in order to construct the temporal interpretation of a sentence. To capture the typological classification of adjuncts in terms of boundary and continuum, the inner AspP node of figure 1 is further branched in the following way:
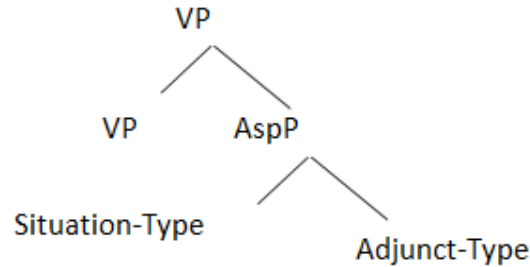
Fig 2: Further Specification of VP-internal Aspectualities

As per this proposal the sole concern of the VP-internal aspect is the situation type, adjunct type and the interaction holding between them; whereas the other VP as the sister node of AspP deals with the argument structure.

5 Conclusion

Integration of syntax and semantics is crucial, since it shows how the semantics of *verb*, *tense*, *aspect* (both grammatical and lexical), and the *adjunct* interact with each other, while construing the temporal interpretation of a sentence by conceptualizing the relevant contextual information of both implicit and explicit types.

References

Arche, M.J. (2006). *Individuals in Time: Tense, Aspect and the Individual/Stage Distinction*, Amsterdam: John Benjamin Publishing Co.

Chomsky, N. (1995). *The Minimalist Program*, Massachusetts: The MIT Press.

Dowty, D. (1979). *Word Meaning and Montague Grammar*, Reidel: Dordrecht.

Reichenbach, H. (1947). *Elements of Symbolic Logic*, New York: The Free Press.

Vendler, Z. (1957). Verbs and Times. *The Philosophical Review*, Vol. LXVI, 143-160.

Verkuyl, H. (1989). Aspectual Classes and Aspectual Composition. *Linguistics and Philosophy*, Vol. 12, 39-94.

Verkuyl, H (1993). *The Theory of Aspectuality: The Interaction between Temporal and Atemporal Structures*, Cambridge: Cambridge University Press.

Biolinguistic Diversity Index of India

Ritesh Kumar
Centre for Linguistics
Jawaharlal Nehru University
New Delhi, India
riteshkrjnu@gmail.com

Abstract

Harmon and Loh calculated the 'Index of Biocultural Diversity' (IBCD) at the global-level (they calculated the index for each country and then compared them). However the problem with their methodology was that it could not be used for smaller areas or intra-country calculation of index. So it required some modification. In this paper I have calculated the IBLD (Index of Biolinguistic Diversity) of India using a similar methodology. However I have introduced some modifications. Instead of taking the politically divided states as the reference point for comparison, I have taken the eco-regions of India as the reference point. I have also calculated the Spearman's rank correlation coefficient between the ranks of the eco-regions on the basis of biodiversity and linguistic diversity so as to see whether these are correlated at the intra-country level (in India, in particular). The conclusions are not exactly at par with the expectations but still the correlation is established.

1 Biocultural diversity: An Introduction

Biocultural Diversity is a very recent concept—not more than a decade old—which is based on the fundamental assumption that there is an inextricable link between biological and cultural and linguistic diversity. It draws from anthropological, ethnobiological, ethnoecological, and linguistic insights regarding the relationships of human language, knowledge, culture and practices with the environment and nature.

Biocultural diversity unifies the diversity of life in all of its manifestations: biological, cultural, and linguistic. These are interrelated and have coevolved within a complex socio-ecological adaptive system. There are three basic assumptions underlying the concept of biocultural diversity.

1 The diversity of life does not only mean the diversity of plants and animal species, habitats, and ecosystems found on the planet, but also the diversity of human cultures and languages.

2 These diversities are not separate or exist in parallel realms, but rather they interact with and affect one another in complex ways.

3 The interaction among these diversities have developed over time through mutual adaptation between humans and the environment at the local level, which is, probably, of a coevolutionary nature and, thus, a causal one (Maffi, 2005, 2007a, 2007b).

The cumulative effect of all these local interlinkages, interdependencies and interaction between the humans and the environment implies that at the global level biodiversity and cultural diversity are also interlinked and interdependent. Thus, it has significant implications for the conservation of both the diversities. Recent global cross-mappings of the distributions of biodiversity and linguistic diversity (taken as a proxy for cultural diversity) have revealed significant geographic overlaps between the two diversities, especially in the tropics. Moreover, they have shown a strong coincidence between biologically and linguistically megadiverse countries (Loh & Harman, 2005). The way in which the local and the global biocultural diversities interact is represented in the following diagram.
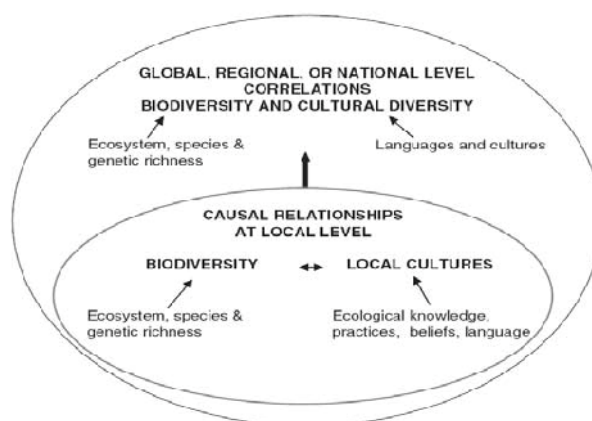
Fig1: Causal relationship at local level and its relation to the diversity at global level.
Source: original figure by Ellen Woodley, 2005

A global cross-mapping of the endemic languages and the endemic higher vertebrate species brings out the remarkable overlap between linguistic and biological diversity throughout the world. Similar results can be obtained by cross-mapping the endemic languages and the flowering plant species. Moreover, comparison of a list of endemic languages with the IUCN list of 'megadiversity' countries reveals that 10 out of the top 12 megadiversity countries (or 83%) also figure among the top 25 countries for endemic languages (Skutnabb-Kangas, 2008).

It has been noted that generally the social factors combine with the geographic and climatic factors leading to a higher or lower diversity. For example, geography and climate of a particular area affects its carrying capacity and access to resources for human use. Ease of access to abundant resources seems to favor localized boundary formation and diversification of larger numbers of small human societies and languages. Where resources are scarce, the necessity to have access to a larger territory to meet subsistence needs favors smaller numbers of widely distributed populations and languages (Carder, 2008). The development of complex societies and large-scale economies, which tend to spread and expand beyond their borders, also correlates with a lowering of both linguistic and biological diversity. Moreover, there is a significant overlap between the location of threatened ecosystems and threatened languages. On the other hand, low population density, at least in tropical areas, seems to correlate positively with high biocultural diversity (Maffi, 1998).

2 IBCD: An Index for Measuring Biocultural Diversity

In order to sustain and protect biocultural diversity we must have a measure of the diversity we have; the diversity we are losing and the rate at which global diversity is decreasing. In 2005, Harmon and Loh developed the blueprint for an Index of Biocultural Diversity (IBCD), whose purpose is to measure the condition and trends in biocultural diversity on a country-to country basis (the level at which the available data sets are organized). Till then linguistic diversity alone was taken as the proxy for cultural diversity. But they took three cultural indicators: languages, ethnicities and religions as these were the aspects of culture for which global data (on a country-to-country level) was available. Finally IBCD is calculated by aggregating data on the three cultural indicators with data on the diversity of bird/mammal species and plant species as indicators for biodiversity (also selected on the basis of global data availability).

The IBCD has three parts:

• A biocultural diversity richness component (BCDRICH), which is a relative measure of a country's 'raw' biocultural diversity using unadjusted counts of the five indicators.

•An areal component (BCD-AREA), which adjusts the indicators for land area and therefore measures a country's biocultural diversity relative to its physical extent.

•A population component (BCD-POP), which adjusts the indicators for human population and therefore measures a country's biocultural diversity relative to its population size.

The IBCD gives equal weight to cultural and biological diversity, so a country's overall biocultural diversity score is calculated as the average of its cultural diversity score (CD) and its biological diversity score (BD).

IBCD = (CD + BD)/2

In measuring a country's cultural diversity CD, equal weight is given to linguistic, religious and ethnic diversity. Therefore CD is calculated as the average of a country's language diversity (LD), religious diversity (RD), and ethnic group diversity (ED):

CD = (LD + RD + ED)/3

In measuring biodiversity BD, equal weight is given to animal species diversity (using birds and mammals as a proxy for all animal species. Marine mammals are excluded from the analysis) and plant species diversity. Therefore BD is calculated as the average of a country's bird and mammal species diversity (MD), and plant species diversity (PD):

BD = (MD + PD)/2

Each indicator is given an equal Weightage as this is the simplest way of calculating the index. As an aggregated index, the IBCD could be calculated using different Weightage, to give greater or lesser importance to any of the five component indicators.

To derive country scores for each of the five component indicators, each country's richness value was compared with the global value. For example, for language diversity, LD is calculated as the log of the number of languages spoken in a country divided by the log of the number of languages spoken worldwide (see the Table above).

LD = (log Li)/ (log L world) where Li is the number of languages spoken in country i, L world is the number of languages spoken in the world (currently 6800).

The calculation was repeated for the other four indicators to derive BCD-RICH.

To compensate for the fact that large countries tend to have a greater biological and cultural diversity than small ones simply because of their greater area (or greater population), two additional diversity values for each country was calculated by adjusting first for land area (BCD-AREA) and second for population size (BCD-POP). This was done by measuring how much more or less diverse a country is in comparison with an expected value based on its area or population alone.

The process was repeated for the other four indicators to derive BCD-AREA and BCD-POP. The expected diversity was calculated using the standard formula for the species–area relationship log S = c + z log A, where

S = number of species,

A = area, and c and z are constants derived from observation.

To calculate the deviation of each country from its expected value, the expected log Ni value was subtracted from the observed log Ni value. The index is calibrated such that the world, or maximum, value is set equal to 1.0, the minimum value is set equal to zero and the average or typical value is 0.5 (meaning no more or less diverse than expected given a country's area or population).

3 Calculating the IBCD within the Ecoregions of India

The first problem in calculating the IBCD within India was to decide upon the reference points within the country. At the global level it was the individual country that formed the reference point. I had two options before me—a) the politically demarcated 'states' of India, which has practically

no scientific basis for their division; b) the eco-regions that the country is divided into by the WWF. Obviously, I chose the latter one as that is more scientifically divided on the basis of biological diversity. An eco-region is a relatively large unit of land that contains geographically distinct assemblage of natural communities with boundaries that approximate the original extent of natural communities prior to the major land use change. Eco-regions:

- share a large majority of their species and ecological dynamics;
- share similar environmental conditions, and;
- interact ecologically in ways that are critical for their long-term persistence. The whole globe is divided into 867 ecoregions and there are the organized data for the biological diversity in these regions.

For the purpose of calculating the IBCD of different ecoregions of India, I have adapted the methodology described above to suit my needs. And the index that I have calculated is not that of 'biocultural' diversity but rather it is something like 'biolinguistic' diversity. I have looked only at the correlation between linguistic diversity and biological diversity in these regions. So it might be rephrased as IBLD instead of IBCD. Moreover I have also not adjusted the data with reference to the area and population. So I have calculated just IBLD-RICH and no IBLD-POP and IBLD-AREA. This has been done because of three reasons:

- My work is basically concerned with the linguistic and biological diversity and so I concentrated on the correlation of these two.
- Lack of proper, authentic data in the field of cultural diversity. For biological diversity I got the data from the WWF website. But for the cultural diversity there was a lot of variations. We can take the example of languages itself. Census 2001 lists just around 122 languages (rest are dialects according to the Census but that is highly contestable); the linguists in India take the total languages to be around 1635 (which is listed as the number of 'rationalised' mother tongues); People of India gives a figure of 325; and Ethnologue lists 414 languages for India. As we can see there are huge differences among different sources and none of these could be called 'unauthentic' or 'unofficial'. I have taken the data of ethnologue for the present study because I think that has an International acceptance. Moreover ethnologue clearly lists the regions in which language is spoken, which proved to be very useful for me. For other two cultural parameters—religions and ethnic groups—the scenario is even more dismal.
- Because of the time constraints. Due to the lack of a proper, organized data for the country, I had to give a lot of time for the collection and compilation of data. For example I got the animal diversity data separately by ecoregions. So I had to find them separately. And then for data on languages, I had to reconstruct the whole data, using the map of the ecoregions of India and the languages of India given in ethnologue. Comparing these two sources, I calculated the number of languages per region. Ethnologue lists the regions within which a particular language in India is spoken and ecoregion is also defined by these regions. This was very time-consuming and the data also seems to be a bit fabricated even though I have tried to remain as close to the original data as possible. I had no other way out; so I had to do it.

Besides the calculation of this IBLD-RICH, given by Harmon and Loh, I have also calculated the Spearman's Rank Correlation Coefficient. It is a simple statistical tool to compare the ranks of two variables and see if they are correlated. It is given by the following formula:

Spearman's Rank Correlation Coefficient ( ) = $[1-\{(6 \quad d^2)/n \ (n^2-1)\}]$ Where d= difference between the ranks of the two variables (for our case it is the difference between the rank for biological and linguistic diversity of each ecoregion) n= Total number of objects for which rank is given (for our case it is the total number of ecoregions)

For the purpose of calculating this rank correlation coefficient, I have ranked the ecoregions separately according to the biological diversity (BD) score and linguistic diversity (LD) score. These two scores are calculated according to the methodology given by Harmon and Loh. Ranking is done in descending order i.e., rank 1 is given to the ecoregion that has the highest score and 37 to the one having the lowest score (total number of ecoregions being 37 in India).

All the significant calculations and findings are given in the tables included in the appendix. Table 1 gives the integrated biolinguistic diversity index; and Table 2 gives the final correlation coefficient of the ranks of biological and linguistic diversity.

4 Analysis and Conclusion

The biolinguistic index calculated by the method described above does not actually give the expected result. At the surface level there does not seem to be any correlation between the biological and linguistic diversity, unlike in the case of the global index calculated by Harmon and Loh (2005). However, the rank correlation coefficient does show some semblance of a correlation. The range of this coefficient is from 0 to 1, where 0 means not correlated and 1 means completely correlated. The correlation coefficient of the ranks of biological and linguistic diversity comes out to be approx. 0.456, which is almost mid-way, although still on the lower side. It proves that they are somehow correlated but this value does not completely support the hypothesis that the two are correlated. Moreover, if we draw a bar diagram (Chart 1) to compare their index scores then we can very clearly see that there is hardly any considerable correlation between the two and there are more exceptions than rules. However with the use of some more sophisticated tools we can see that there do exist some kind of correlation. Chart 2 is a scatter diagram where we plot the value of the indices on the x-axis and on the y-axis the ranks of the ecoregion are plotted. In the diagram that I have shown, the ranks are decided by arranging the indices of the linguistic diversity in descending order. So at the place of rank 1, we plot the value of both the biological and linguistic diversity index of the ecoregion concerned and similarly we plot the whole of the scatter diagram. Then all the scattered points of both the data are joined one by one separately. And finally a straight line is drawn through the points such that it passes through the maximum number of points. This straight line shows the 'linear regression' and it is called 'line of regression' and the rest of the points are chiefly scattered around it. Literally, 'regression' means 'stepping back towards the average'. It is a simple statistical tool which gives the measure of the average relationship between two or more variables in terms of the original unit of the data. In this case different index scores of different ecoregions are our variables and these variables are to be compared. For both biological and linguistic diversity, I have made different lines of regression on the same sheet. Now if we compare the two lines of regression we see that in both the cases the average tendency is a decrease in the index value as we move towards lower rank; so this shows the correlational pattern. However the differences are too stark to be neglected. In the case of biological diversity there is a huge amount of scattering the standard deviation from the average. Moreover from rank 1 to rank 37 there is nothing like a very significant or sharp fall in the average index score. But on the other hand in the case of linguistic diversity the scattering around the line of regression is not much (it should be so as we have arranged these indices in descending order) and there is a very considerable difference between the average score of ecoregion ranked 1 and ecoregion ranked 37.

These discrepancies in the final result can be accorded to two main reasons:

- Absence of very authentic data: The data on both plant diversity and linguistic diversity is not very reliable and their authenticity is always on the questionable side.
- Problems in the method of calculation: Harmon and Loh themselves had pointed out that the index calculated by their method may not prove to be very useful in case of intra-country

scenarios. Moreover these indices only give general conditions and they are not expected to give an in-depth analysis. And in this case, as shown in the scatter diagram by the line of regression, I think that it has indeed pointed towards the general trend which is in consonance with the hypothesis that biological and linguistic diversity are somehow correlated.

In order to make this index more powerful, efforts are now underway to develop time-series data (data calculated over a large period of time) on linguistic diversity, as well as the methodology for a locally appropriate, globally applicable indicator directly focused on trends of retention or loss of traditional environmental knowledge over time. Both of these indicators might contribute to the needs of the linguistic, cultural and biological diversity, and will help in finding an answer to one of the basic questions in the field of biocultural diversity: is the world's cultural diversity indeed in decline, and, if so, how fast? Correlated with time-series data on biodiversity, these new indicators will also show whether trends in cultural diversity and biodiversity mirror each other.

Also, an expert group on language endangerment and language maintenance gathered by UNESCO has put forth a set of recommendations for the assessment of linguistic vitality that is also expected to provide some help for the development of linguistic diversity indicators. They point out that sheer trends in "language richness" (number of different languages) are not a fully adequate indicator of the state of languages. Better data on numbers of speakers over time and other sociolinguistic "vital statistics," particularly on intergenerational language transmission, contexts of use, availability of mother tongue education, and so forth, will be needed for this purpose. A methodology has recently been developed for testing linguistic vitality at the local level and identifying the factors (such as age, gender, special roles, etc.) that affect linguistic ability.

This study of the linguistic and biological diversity within India has reaffirmed that these two (and consequently cultural diversity also) are correlated. But at the same time it has also reaffirmed the need for much better database and also some improvement in the methodology for measuring these correlations as well as the trends.

References

Carder, Maurice. 2008. *A Word of Difference*, from 'Resurgence Magazine', 250.

Gupta, S.C. and V.K. Kapoor. 2003. *Fundamentals of Mathematical Statistics*. Sultan Chand & Sons.

Loh, Jonathan and David Harmon. 2005. *A global index of biocultural diversity*, from 'Ecological Indicators'.

Maffi, Luisa. 1998. *Language: A Resource for Nature*, from 'The UNESCO Journal on the Environment and National Resources Research', 34(4).

Maffi, Luisa. 2005. *Linguistic, Cultural and Biological Diversity*, from 'Annual Review of Anthropology' 2005.

Maffi, Luisa. 2007a. *Biocultural Diversity and Sustainability* 'The Sage Handbook of Environment and Society'.

Maffi, Luisa. 2007b. *Cultural Vitality*, from 'Resurgence Magazine', 250.

Skutnabb-Kangas, Tove. 2008. *Linguistic Genocide in Education or Worldwide Diversity and Human Rights*. Orient Longman.

Table 1
Biolinguistic Diversity Index

Eco_id	Eco_code	Eco_regions	Biolinguistic Diversity Index
40101	IM0101	Andaman Islands rain forests	0.5429
40105	IM0105	Brahmaputra Valley semi-evergreen forests	0.7908
40111	IM0111	Eastern highlands moist deciduous forests	0.757
40115	IM0115	Himalayan subtropical broadleaf forests	0.6254
40120	IM0120	Lower Gangetic Plains moist deciduous forests	0.8013
40124	IM0124	Malabar Coast moist forests	0.5903
40125	IM0125	Maldives-Lakshadweep-Chagos Archipelago tropical moist forests	NA
40126	IM0126	Meghalaya subtropical forests	0.7552
40131	IM0131	Mizoram-Manipur-Kachin rain forests	0.7465
40133	IM0133	Nicobar Islands rain forests	0.5478
40134	IM0134	North Western Ghats moist deciduous forests	0.6659
40135	IM0135	North Western Ghats montane rain forests	0.6772
40142	IM0142	Orissa semi-evergreen forests	0.5925
40150	IM0150	South Western Ghats moist deciduous forests	0.6535
40151	IM0151	South Western Ghats montane rain forests	0.6717
40162	IM0162	Sundarbans freshwater swamp forests	0.5496
40166	IM0166	Upper Gangetic Plains moist deciduous forests	0.6435
40201	IM0201	Central Deccan Plateau dry deciduous forests	0.6416
40203	IM0203	Chhota-Nagpur dry deciduous forests	0.7154
40204	IM0204	East Deccan dry-evergreen forests	0.5983
40206	IM0206	Khathiar-Gir dry deciduous forests	0.6315
40207	IM0207	Narmada Valley dry deciduous forests	0.6817
40208	IM0208	Northern dry deciduous forests	0.6577
40209	IM0209	South Deccan Plateau dry deciduous forests	0.6828
40301	IM0301	Himalayan subtropical pine forests	0.6805
40303	IM0303	Northeast India-Myanmar pine forests	0.6263
40401	IM0401	Eastern Himalayan broadleaf forests	0.7383
40403	IM0403	Western Himalayan broadleaf forests	0.6087
40501	IM0501	Eastern Himalayan subalpine conifer forests	0.5695
40502	IM0502	Western Himalayan subalpine conifer forests	0.5682
40901	IM0901	Rann of Kutch seasonal salt marsh	0.5774
41301	IM1301	Deccan thorn scrub forests	0.7479
41303	IM1303	Northwestern thorn scrub forests	0.7474
41304	IM1304	Thar desert	0.6155

41401	IM1401	Godavari-Krishna mangroves	NA
41406	IM1406	Sundarbans mangroves	NA
81003	PA1003	Eastern Himalayan alpine shrub and meadows	0.50225
81006	PA1006	Karakoram-West Tibetan Plateau alpine steppe	0.5904
81012	PA1012	Northwestern Himalayan alpine shrub and meadows	0.65285
81021	PA1021	Western Himalayan alpine shrub and Meadows	0.6074

Table 2
Calculation of Rank Correlation Coefficient

eco-regions	biodiversity rank	linguistic diversity rank	Difference in the rank	
	(br)	(lr)	d=br-lr	d^2
Andaman Islands rain forests	35	32	3	9
Brahmaputra Valley semi-evergreen forests	6	1	5	25
Eastern highlands moist deciduous forests	18	3	15	225
Himalayan subtropical broadleaf forests	14	25	-11	121
Lower Gangetic Plains moist deciduous forests	3	2	1	1
Malabar Coast moist forests	12	33	-21	441
Meghalaya subtropical forests	4	7	-3	9
Mizoram-Manipur-Kachin rain forests	1	8	-7	49
Nicobar Islands rain forests	36	31	5	25
North Western Ghats moist deciduous forests	9	17	-8	64
North Western Ghats montane rain forests	7	14	-7	49
Orissa semi-evergreen forests	26	30	-4	16
South Western Ghats moist deciduous forests	8	19	-11	121
South Western Ghats montane rain forests	5	18	-13	169
Sundarbans freshwater swamp forests	33	34	-1	1
Upper Gangetic Plains moist deciduous forests	11	22	-11	121
Central Deccan Plateau dry deciduous forests	16	20	-4	16
Chhota-Nagpur dry deciduous forests	24	6	18	324

73

East Deccan dry-evergreen forests	28	26	2	4
Khathiar-Gir dry deciduous forests	22	23	-1	1
Narmada Valley dry deciduous forests	29	9	20	400
Northern dry deciduous forests	20	15	5	25
South Deccan Plateau dry deciduous forests	15	11	4	16
Himalayan subtropical pine forests	13	12	1	1
Northeast India-Myanmar pine forests	25	21	4	16
Eastern Himalayan broadleaf forests	2	10	-8	64
Western Himalayan broadleaf forests	32	24	8	64
Eastern Himalayan subalpine conifer forests	21	36	-15	225
Western Himalayan subalpine conifer forests	27	35	-8	64
Rann of Kutch seasonal salt marsh	34	27	7	49
Deccan thorn scrub forests	19	4	15	225
Northwestern thorn scrub forests	17	5	12	144
Thar desert	37	16	21	441
Eastern Himalayan alpine shrub and meadows	10	37	-27	729
Karakoram-West Tibetan Plateau alpine steppe	31	28	3	9
Northwestern Himalayan alpine shrub and meadows	30	13	17	289
Western Himalayan alpine shrub and Meadows	23	29	-6	36
	Total	**0**	**4588**

BADCORRELATION	10	27.03%
V.BADCORRELATION	4	10.81%
SOMEWHAT CORRELATED	23	62.16%
Rank Correlation Coefficient=	1-[6Σ(d^2)/n(n^2-1)] 0.456140351	

Chart 1: Comparison of Biodiversity and Linguistic Diversity in Different Ecoregions of India
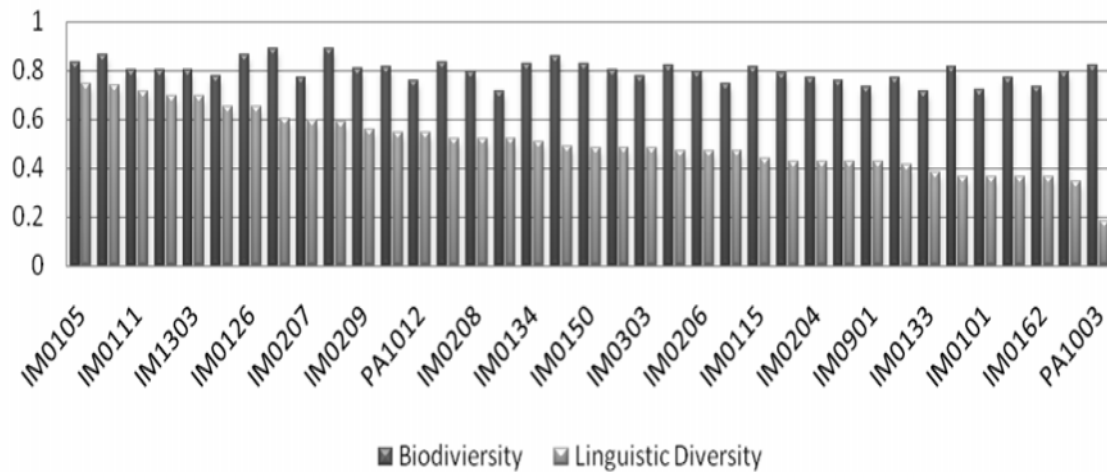
Chart 2: Scatter Diagram Showing the Comparative Trends of Biodiversity and Linguistic Diversity
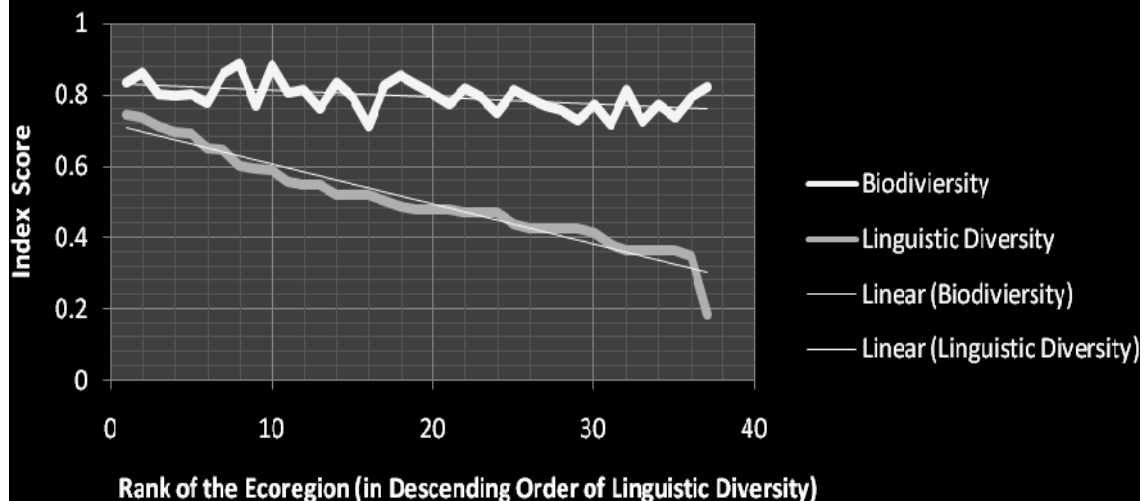
Gilchrist's 'A Grammar of Hindoostanee Language': Some colonial and contemporary imprints

Santosh Kumar
University of Delhi, Delhi, INDIA
santoshk.du@email.com

Abstract

The contextual and textual analysis of a piece of grammar has always been an interesting and challenging area of inquiry**.** This paper examines Gilchrist's *A Grammar of Hindoostanee Language* with the aim to show that often grammar-writing can be seen as a platform to provide a framework to 'keep in circulation' the imperialist project (Bhattacharya 2004) The history of writing grammars shows that grammars are written for specific linguistic reasons such as to provide a principled description of a particular language; to be a basis for language pedagogy; to compare one grammar with another for typological, historical, and aerial characteristics or for pedagogically oriented contrastive analysis, and to test linguistic theories (Ferguson, 1978). However, there are some non-linguistic reasons such as culture, religion, ecology, aesthetics, pragmatism, etc. that influence the practice of writing grammar, wittingly or otherwise (Scharfe, 1977). The latter perspective provides a significant site for raising questions of representation, power and historicity. John Borthwick Gilchrist's '*A Grammar of Hindoostanee language or part third of volume first of a system of Hindoostanee philology*' (Calcutta, Chronicle Press, 1796) is one of the earliest grammars written on the Hindustani language. This well-studied text can however be seen as a representative of the colonial language policies put in place by the British to serve their needs in addition to its being representative of the British attitudes towards Indian vernaculars in general and Hindustani in particular. Notwithstanding problematic aspects such as these in a purportedly scientific activity like grammar writing, this paper also presents comparative evidence to show that the contemporary mindset is not without blemish when it comes to pursuing an imperialist agenda through academic writing.

1 Objective

The paper presents certain excerpts from John Borthwick Gilchrist's well known grammar, *A Grammar of Hindoostanee Language or part third of volume first of a system of Hindoostanee Philology* published in 1796, in order to show how this grammar and the colonial language policies influenced each other.

2 Introduction

The contextual and textual analysis of a piece of grammar has always been an interesting and challenging area of inquiry**.** This paper examines Gilchrist's *A Grammar of Hindoostanee Language* with the aim of showing that often grammar-writing can be seen to provide a framework to 'keep in circulation' the imperialist project (Bhattacharya 2004). The history of writing grammars shows that grammars are written for specific linguistic reasons such as to provide a principled description of a particular language; to be a basis for language pedagogy; to compare one grammar with another for typological, historical, and aerial characteristics or for pedagogically oriented contrastive analysis, and to test linguistic theories (Ferguson, 1978). However, there are some non-linguistic reasons such as culture, religion, ecology, aesthetics, pragmatism, etc. that influence the

practice of writing grammar, wittingly or otherwise (Scharfe, 1977). The latter perspective provides a significant site for raising questions of representation, power and historicity. John Borthwick Gilchrist's '*A Grammar of Hindoostanee language or part third of volume first of a system of Hindoostanee philology*' (Calcutta, Chronicle Press, 1796) is one of the earliest grammars written on the Hindustani language. Since, Hindi grammatical tradition is representative of the vernacular tradition and its form is often governed by non-linguistic factors like social, political, attitudinal, etc. This well-studied text can however be seen as a representative of the colonial language policies put in place by the British to serve their needs in addition to its being representative of the British attitudes towards Indian vernaculars in general and Hindustani in particular. Notwithstanding problematic issues such as these in a purportedly scientific activity like grammar writing, this paper also presents comparative evidence to show that the contemporary mindset is not without blemish when it comes to pursuing an imperialist agenda through academic writing. So, an attempt has been taken to utilize the rare and invaluable sociolinguistic information about language attitude, linguistic variations and data selection stored in this grammar. Gilchrist begins with stating that the grammar was written for the British subject only.

> "*The present work being expressly designed for the improvement and advancement of British subjects only, a general or continental pronunciation has been little attended to; such foreigners, however, as may honour the orthographical table and notes with a careful perusal, will perceive that they are by no means neglect;*" Gilchrist (1796: 3).

3 Orientalism and Philology

It is quite likely that Edward Said, perhaps, was not aware of Gilchrist's work and indentified instead *Tableau de gènèrale* of Silvestre de Sacy (1826) as a model of orientalist world view *vis-a-vis* language. In addition to it, Said was interested in grammars and lexicons dealing with Arabic, he found Sacy's work to be a model that later grammarians followed. Given that Gilchrist's work pre-dates Sacy's, it is likely that a re-working of the details of the *Tableau* will reveal attentive world views in terms of grammar construction.

Philology according to Said equips the philologist with the power to see himself as the centre-piece, the originator of knowledge. Said cites Renan (1890) for having religion in his mind while he described philology- for example, what philology, like religion, teaches us is about the origins of humanity, civilization and language,' (Said, 1978). According to Ernest Renan, an Oriental philologist enables a general view of human life and of the system of things: "*Me, being there at the centre, inhaling the perfume of everything, judging, comparing, combining, inducing – in this way I shall arrive at the very system of things.*" (*L'Avenir de la Science* 1890, as quoted in Bhattacharya, 2004).

4 Colonial language policies and Hindi grammar

It is believed that Hindi or Hindustani language evolved in the camps of soldiers during the colonial period as a means of interaction among the higher British officers and native soldiers from different linguistic backgrounds and it was here that Hindi or Hindustani was born. Hindi grammar is a by-product of the colonial era and was born shortly after the arrival of Europeans in India. According to the renowned Hindi grammarian Guru (1920), the pioneers of Hindi grammar were the Britishers who felt the need to learn the language (Hindi) in the beginning of the 19th century (Bhatia, 1987). The history of Hindi grammatical tradition began with Ketelaar's Original Manuscript (1698), a non-native Dutch scholar. The forward of this grammar is interesting as the copyist abandons his traditional role of commenting on the text and its author rather highlights the

underlying motivations for foreign language (Bhatia, T. 1987). The observations provide a rare insight into the attitude of seventeen-century Europeans toward the learning of second language. Gilchrist echoed the same in his Preface to 'The Stranger's East-Indian Guide to the Hindoostanee language (called Moors), Calcutta, 1802 about the importance and necessity of learning Hindustani.

> *"We must admit, that however learned our British judges, civil and military officers, might otherwise be, they would all prove nowadays wholly unfit for their respective offices in our country, without an adequate knowledge of their mother tongue; and the Hindoostanee being exactly to India, what the English language is colloquially to the U.K., or what the Turkish is to that empire, the inference is so very plain, that he who runs may read."* Guide: Preface, p. xix.

Europeans came to India with two specific goals in mind: commerce and the spread of Christianity. As they started achieving these goals, they were also inspired to establish their administration across India. Communication was one of the major problems in translating these goals into reality. In the absence of a link language between the British and the Indians, it was imperative for the former to break the ice by gaining some knowledge of widely used Indian languages. Apart from Sanskrit and Persian, they preferred to learn Hindi or Hindustani over other languages because it was the language that was understood by a large number of people. Sanskrit and Persian were the languages of the elite; the language of the masses, particularly in North India, was Hindustani. Gilchrist also observed that it was only Hindustani that served as the most important communicable tool for Britishers as it was the language of the people.

> *"Nearly from cape comorin to Kabul, a country about 2000 miles in length and 1400 in breadth within the Ganges, few persons will be found in any large village or town, which have ever been conquered or much frequented by Musalmans, who are not sufficiently conversant in the Hindustan and in many places far beyond the Ganges, even on the source of eastern isles, this language (Hindustani) is current and familiar, or so well known that it can be readily understood."*-Ibid, p. xii

The British arrived on the Indian shores relatively late but they stayed here for a considerable time. Before them, the Portuguese and the Dutch made their appearance while the French and the Danish did so later, but neither of them was able to consolidate their position on the Indian subcontinent. The British, however, took about a hundred year to consolidate their position and acquire a certain level of stability partly because of their language policies. Even though it is true that the British rule at the end of the day was held responsible for the stigmatisation of 'the Indian vernaculars', it is equally true that in the beginning, in order to promote their trade and administrative interests, and also to establish a good relation with the Indian elite, they learnt and promoted Indian languages.

> *"It may be observed that many centuries have not elapsed since the learning, laws, and religion of our own country were preserved, communicated and studied in more than one foreign language, those languages, however, have been fortunately superseded by the English, a sufficient proof that the current language of the country should always be deemed most worthy of public cultivation and private study."* Ibid. xiii.

It would be interesting to examine to what extent Gilchrist's grammar was influenced by the language policies, which enabled the British to anchor here for so long. East India Company (EIC) was formed in 1601 in England. Its arrival on the Indian soil was at Surat, in the early seventeenth century and it marked the beginning of their trade mission. By the beginning of the 18th century, the British were able to establish themselves in India and they succeeded in establishing their factories in Balasore, Madras, Bombay**,** and finally in Calcutta in 1690. The reasons behind the relatively long stay of the British in India than the other Europeans are varied and are of particular importance to understand the colonial mindset. The British were different from other Europeans in at least two

aspects. Their organizational skill and focused policies were better than the others. Not only this, the British had a sense of openness to the new culture they came in contact with. India's dress and food habits, culture, language and social etiquette, etc., did not bother them much and they did not disturb the social and cultural fabric of India. Instead, they appreciated the diversity of India's culture, languages, literatures, arts and crafts, its people and places in the beginning. This can be well understood by D. Forbes who wrote Hindustani Manual in 1845 and in the following year, A Grammar of the Hindustani language (1846) was published and the very popular Linguistic Survey of India by Grierson, G.A 1898. Here, he wrote many descriptions of the Hindustani language.

> "*The grammar of any language is to be learned only through the language and language by means of the grammar, but to learn or rather to attempt to learn, the one without the other, is about as profitable a pursuit as the manufacturing of bricks from straw without clay or from clay without straw.*"-'altera aterius auxilio eget', Forbes, D. 1845. 'The Hindustani Manual,. Preface, p. 7.

This kind of policy shows that the cultural domination works by consent and can (and often does) precede conquest by force. As Viswanathan (1989) quoting Gramsci points out that power produces a situation where the supremacy of social groups manifests itself in two ways, namely, 'domination' and 'intellectual and moral leadership'. She also adds "*that there can, and indeed must be hegemonic activity even before the rise to power, and that one should not count only on the material force which power gives in order to exercise an effective leadership.*" (Viswanathan 1989:1)

The case of colonial India, subject to the vagaries of particular circumstances, is a perfect instance of Gramscian notion. Viswanathan (1989) believes that colonial India's chequered history of cultural confrontation conferred a sense of urgency to voluntary cultural assimilation as the most effective form of political action. She (Viswanathan, 1989) cites J. Farish' minute issued in the Bombay Presidency- "*The Natives must either be kept down by a sense of our power, or they must willingly submit from a conviction that we are more wise, more just, more humane, and more anxious to improve their condition than any other rulers they could possibly have.*"[1]To establish and to maintain the established colonial power, the British Empire in the early years of their arrival have witnessed a wide-ranging, multi-disciplinary debate about the role of cultural activity, and particularly the 'making of knowledge'. Given this interest in the political dimensions of intellectual activity, it is not surprising that a certain amount of attention has been paid to colonial grammars in general and the work of Gilchrist in particular who emphasized the need to learn local language in order to negotiate the minefield of Indian etiquette.

If the subject now treated of can never arrive at perfection, a spirit of exertion, and enquiry, becomes the more requisite, for its progressive improvement, to the *ne plus ultra,* that may soon be reasonably expected from a united efforts of our indefatigable countrymen in this extensive empires won by their valour, supported by their wisdom, and which will most likely be better managed, and longer preserved by our becoming every day more intimately acquainted with the languages, laws, religions, manners, policy, and interests, of its innumerable and multifarious inhabitants. (Gilchrist, 1796: 3)

The efforts to establish Hindustani language as a medium of communication by Gilchrist is an example of how he charged his precedents who called this language as Moors' language, and in a way, influenced colonial language policies. As Gilchrist put it later in the Preface of East-India Guide (1802: viii)

> "*The language which was long degraded under the name of Moors (as a jargon) but at present best known as the Hindustan, is also frequently denominated Hindi, Urdu and*

---

[1]   Minute by J. Farish dated August 28, 1838, quoted in Boman-Behram, B. K. 1943. Educational Controversies in India: The Cultural Conquest of India Under British Imperialism. Bombay: D.B. Taraporevala Sons & Co

Rekhti. It is compounded of the Arabic, Persian and Sanskrit or Bhakha, which last appears to have been in former ages of the current language of Hindustan."

Gilchrist observed that the Hindustani language was the medium of communication among the non-natives who settled in India. As he states in his Guide-

It is moreover the common medium by which the natives in general and many persons of various nations settled in hindoostan, communicate their wants and ideas orally to each other. Of the truth of this, indeed, we ourselves are an evidence, as are the Portuguese, Dutch, French, Danes, Arabs, Turks, Greeks, Armenians, Georgians, Persians, moguls, and Chinese, who often converse together in the hindoostanee, as a more convenient lingua franca for India, than their own individual language. ibid-xi

Gilchrist also suggested the knowledge of Hindustani is the most requisite qualification for the persons who were to land in India.

"the Hindoostanee is no less indispensable, than a knowledge of English evidently is to any stranger residing among, and transacting business with the inhabitants of the united kingdom; and consequently should be treated as the most requisite qualification for the gentlemen bound to the east indies;" ibid-xvii-viii.

5 Conclusion

In this paper I have attempted to show how the attempt to write the grammar of vernacular language called Hindustani by John Borthwick Gilchrist, a military surgeon turned orientalist who wrote '*A Grammar of the Hindoostanee language*' in the last half of the eighteenth century and the first half of the nineteenth century when British were just started to gain some ground on India soil helped the East India Company to achieve its colonial goals. No doubt, this grammar also provide the framework to 'keep in circulation' imperialist project (Bhattacharya 2004), also it became the model for the later grammarians. Thus, the arguments which he put forward in favour of Hindustani language can be seen as the colonial imprints which shaped the colonial language policies.

References

Agnihotri, R.K. and Khanna, A.L. 1997. *Problematizing English in India*. New Delhi: Sage Publications.

Bhattacharya, Tanmoy. 2004. *Hand me my slippers and other such phrases as a part of grammar: Pettigrew's Tangkhul Naga Grammar.* Paper presented at the 26th AICL Meeting, NEHU.

Bhatia, Tej K. 1987. *A History of the Hindi Grammatical Tradition: Hindi-Hindustani Grammar, Grammarians, History and Problem.* The Netherlands: E. J. Brill, Leiden.

Chaudhary, Anant. 1972 *Hindi Vyakaran ka Itihaas*. Patna: Bihar Hindi Granth Akademy.

Cohn, B.S. 1985. *The Command of language and the language of command, In Guha, Ranjeet (ed) 1985 Subaltern Studies IV: Writings on South Asian History and Society*. Delhi pp. 276-329.

Ferguson, C.A. 1978. *Multilingualism as object of linguistic description*, In Kacharu, Braj B (ed) 1978. Linguistics in the Seventies: Directions and Prospects, Department of Linguistics, Urbana, University of Illinois.

Gilchrist, J.B. 1796. *A Grammar of Hindoostanee language or Part third of volume first of a system of Hindoostanee philology*. Culcutta: Chronical Press.

Gilchrist, J.B. 1802a. *The Stranger's East Indian Guide to the Hindoostanee; or Grand Popular Language of India (Improperly Called Moors)*. Culcutta: Hindoostanee Press.

Hussein, Abdirahman A. 2002. *Edward Said: criticism and society*. London and New York: Verso.

Kidwai, S. R. 1972. *Gilchrist and the 'Language of Hindoostan'*. Ph.D. thesis, University of Delhi. Delhi: Rachna Prakashan.

Krishnaswamy, N. and Brude, A.S. 2002. *Review of the Politics of Indians' English: Linguistic Colonialism and the expanding English empire'*. English World-Wide, 21, 151-156.

Niranjana, T. 1993. *Siting Translation: History, Post-structuralism and the colonial context*. Berkeley: University of Californial Press.

Said, Edward W. 1978. *Orientalism*. Routledge and Kegan Paul Ltd.

Spolsky, Bernard 2004. *Language Policy*. Cambridge: Cambridge University Press.

Steadman-Jones, J. 2003. *Richardson's Barometer: Colonial Representation in Grammatical Texts*. In Glaisyer, N and Pennell, S. (ed) 2003, Didactic Literature in England, 1500-1800: Expertise Constructed. Aldershot, Hamshire England: Ashgate Publishing. Ltd.

Steadman-Jones, J. 2007. *Colonialism and Grammatical representation: John Gilchrist and the analysis of the Hindustani language in the late eighteenth and early nineteenth centuries*. Publications of the Philological Society, 41. Blackwell Publishing

Scharfe, Hartmut. 1977. *Grammatical Literature*. Wiesbaden, Otto Harrassowitz.

Suleri, S. 1992. *The Rhetoric of English India*. Chicago: The University of Chicago Press.

Viswanathan, G. 1989. Mask *of Conquest: Literary Studies and British Rule in India*. Columbia: Columbia University Press.

Bhojpuri 'waalaa'

Shailendra Kumar   Neha Vashistha
Department of Linguistics, Banaras Hindu University, Varanasi, INDIA

Abstract

In many north Indian languages, waalaa plays a very diverse functional role. In major languages such as Hindi (Bhatia 2008, Sharma 1958, Shapiro 1989) the topic has been studied in detail. *Waalaa* seems to have more or less similar roles in many of the languages in which it is used and can be categorized on similar ground. However, the uses of waalaa in Bhojpuri deserve a separate discussion. In this paper, we attempt to do this. First of all we examine the various uses of waalaa in Bhojpuri. For this we rely on my own native language intuition and also discuss some relevant data from friends and other native speakers. Secondly, we attempt to make a categorization of waalaa according to its functional roles. In this paper, we examine the various uses of waalaa in Bhojpuri and categorize them according to their various functions. We also examine the constructions that do not allow the uses of waalaa in certain contexts and the constraints on them.

1 Introduction

The particle *waalaa* occurs in Hindi in multiple functional roles and has been studied in the Hindi grammar books and linguistics literature from various points of view (Guru 1925, Sharma 1958, Shapiro 1989, Bhatia 2008). The occurrence of *waalaa* particle is attested in many other Indic languages such as Bhojpuri, Maithili, Magahi, etc. However, there has not been any substantial study on the grammatical property of *waalaa* particle in these languages. It has been generally assumed (without proper study) that the grammatical property of *waalaa* in these languages is identical or similar to those in Hindi. However, it is necessary that each language is examined in its own right for the distribution of *waalaa* attested in it. This paper is an initial attempt to examine the distribution of *waalaa* in Bhojpuri. On the basis of this, the paper also attempts to identify the different functional roles that *waalaa* particle has in Bhojpuri and categorize them according to its various functions. For instance, the use of *waalaa* in sentence (1a) and (1b) are different and need to be categorized accordingly.

(1) a. *tohaar waalaa kaam kab hoi? (sabhane ka to kaam kabka hogawaa baa).*
         your PRT[1] work when happen-will (all of PRT work long time happened)
         'When will YOUR work happen?' (all the others have finished their work long time back)
     b. *phon-waalaa pEsaa kahaaN baa?*
         phone-for/from   money where is
         'Where is the money meant for/that came from phone?'

In (1a), *waalaa* is used as a contrastive focus marker as it can be contrasted with the work assigned to others (this is shown in the bracketed part of the example). In (1b), on the other hand, the use of *waalaa* can have many interpretations as is shown in the English translation of the sentence.

---

[1]   PRT: Particle, m: masculine, f: feminine, sg: singular, pl: plural,

Bhojpuri is spoken in a large territory spreading across UP and Bihar and exhibits variations at all the levels of grammar across different geographical areas. In this paper, we discuss examples of Bhojpuri spoken in the Mirzapur district of UP. As mentioned, variations are noticed across the different varieties of Bhojpuri (for instance, Banarasi variety is different from Mirzapuri variety).

In section 2, we discuss the distributional properties of *waalaa* by examining its different compatible positions with other grammatical categories. In section 3, we discuss their various functional roles and show how multiple interpretations of *waalaa* are conditioned. In section 4, we conclude the paper.

2 Distribution of *waalaa* in Bhojpuri

The distribution of *waalaa* is spread across grammatical categories in Bhojpuri. It can be seen occurring with nouns, verbs, adjectives, adverbs, etc. W*aalaa* inflects for gender and number of the noun that follows it and occurs in the following inflected forms:
- waalaa (m.sg)
  E.g. *paRe waalaa laikaa* 'studious boy'
- waalii (f.sg)
  E.g. *paRe waalii laikii* 'studious girl'
- waalan/waale (m.pl)
  E.g. *paRe waalan laikan* 'studious boys'
- waalin (f.pl)
  E.g. *paRe waalin laikian* 'studious girls'

Below we look at the instances of *waalaa* occurring with different grammatical categories:

2.1 Noun-*waalaa*
Occurrence of *waalaa* is the most productive in case of nouns. It can be attached with any kind of nominal element in the configuration of (N$_1$-*waalaa* N$_2$).

N$_1$
Proper names
Person's name: *raam waalii kitabawaa* 'Ram's book'
Place names: *dilli waalan laikan* 'Boys from Delhi'
Institution's name: *jnu waalan laikan* 'Boys from JNU'
Common names: *laikii waalii batiyaa*
                          girl   PRT   matter
                          'matters pertaining to girls'
Locative: *shahar waalii batiyaa*
                          city   PRT   matter
                          'matters pertaining to cities'
                          khet waalii balTiyaa
                          field PRT   bucket
                          'bucket for/which is used in a field'
                          rasoii waalii balTiyaa
                          kitchen PRT   bucket
                          'bucket for kitchen'

2.2 Pronoun-*waalaa*

Occurrence of *waalaa* is also attested with pronominal elements and demonstratives. Some representative examples are given below.

Personal Pronoun: *aap waalii batiyaa*
                       you PRT   matter
                       'the matter that you said

Possessive pronoun: *okar waalaa sujhaav*
                       his   PRT   suggestion
                       'the suggestion that he gave'

Demonstrative: *u waalii batiyaa*
                       that PRT matter
                       'that matter'

2.3 Adverbial-*waalaa*

waalaa also occurs with all the types of adverbial elements. Examples are given below:

Temporal: *kalhi waalii ghaTanaa*
                 yesterday PRT event
                 'the incident that happened yesterday'
                 ratiyaa waalii batiyaa
                 night PRT   matter
                 'the matter pertaining to night/the matter that took place in the (last) night'

Locative: *upar waalaa kamarawaa*
                 upstair PRT room
                 'the room located upstairs'
                 bagale waalaa gharwaa (kekar ha)
                 side   PRT   house
                 'the house located nearby'
                 ehar-ohar waalaa (mat li aaya)
                 here-there PRT(not take bring)
                 '(do not bring (things)) from here and there'

2.4 Verb-*waalaa*

waalaa is also attached with verbs either in gerundive form or in participial form. We can look at the representative examples below:

Gerundive form: *jaay waalan laikan*
                  going PRT boys
                  'the boys who are (about) to go'
                  cale waalii mashiniyaa
                  working PRT machine
                  'the machine that works'

Aspectual form: *u taraf jaat waalan laikan*
                  that side going PRT boys
                  'the boys who are going that side'
                  uhaaN baiThii waalii laikii
                  there sitting PRT girl
                  'the girls who are sitting there'

2.5 Adjectives

waalaa occurs with all types of adjectives quite frequently and gives different interpretations. Howevr, *waalaa* does not occur with derived adjectives[2]. Some of the examples are given below:

Color:	*niillii waalii busharT*
	blue PRT shirt
	'the blue (bu)shirt'
	kalii waalii gaay
	black PRT cow
	'the black cow'
Shape/size:	*baRaa waalaa ghar*
	big PRT house
	'the big house'
	gol waalii Tophii
	round PRT cap
	'the round-shaped cap'
Quality:	*baDiyaa waalaa bhojan*
	good PRT food
	'the good food'
Derived adjectives:	**rashtriya waalaa giit*
	national PRT song
	'national song'

Postpositions do not take *waalaa* (since *waalaa* itself occurs as a postposition in N_1-*waalaa*-N_2 constructions). The occurrence of *waalaa* with *upar* 'up', *niice* 'below' like elements are instances of *waalaa* with adverbials since in these uses of *upar* 'up' and *niice* 'below' are adverbial.

3 Functions of *waalaa*

In this section, we discuss the various functional roles in that *waalaa* take with different grammatical categories. In the following discussion we list these various functional roles of *waalaa* in Bhojpuri.

3.1 Possessive/genitive function

In N_1-*waalaa*-N_2 distribution, one of the most common functions of *waalaa* is to denote possession relation. However, many more relations can potentially be assigned in these cases. For instance, in example (1) below *raam* 'Ram' can be assigned roles like agent (author), seller (of the book), etc.

(1)   *raam waalii kitabiyaa*
          Ram PRT book
          'Ram's book'

3.2 Temporal function

waalaa also denote temporal function, particularly in the case where it is used with a verb in gerundive form. In this case, *waalaa* denotes future or probability marker of time, as in (2).

(2)   *raam uhaaN jaay waalaa ha.*
          Ram there going PRT is
          'Ram is about to go there.'

---

[2]   Details need to be worked out with more varieties of examples.

3.3 Locative function

waalaa can also function as locative marker, particularly when it occurs with location denoting words, as in (3).

(3)   *amerikaa waalii laikii*
         America PRT   girl
         'the girl from America'

3.4 Modifier function

waalaa functions as a modifier to noun. Generally it occurs in typical relative clause-like modifier function.

(4)   a.   *ohar jaay waalaa laikaa*
           that-side going PRT boys
           'the boys who are going that side'
      b.   *ohaaN baiThii waalii laikii*
           there sitting   PRT girls
           'the girls who are sitting there'

3.5 Agentive function

waalaa marks agentive function particularly when it occurs with a verb in gerundive form preceded by a (common) noun, as in (5).

(5)   *sabjii bece waale (log)*
         vegetables selling PRT (people)
         'vegetable sellers'

The functional roles of *waalaa* listed above are some of its most common uses. The data used in this paper have been collected in informal settings, mostly the spoken form of Bhojpuri. For a detailed discussion and a comprehensive listing of the functional roles of *waalaa* in Bhojpuri, we need a large corpus which is not available at present.

4 Concluding Remarks

In this work, we have made an initial attempt to examine the *waalaa* particle in Bhojpuri with a view to describe its distributional property and functional categorization. In our observation, *waalaa* has more varied and productive roles in Bhojpuri than that in Hindi. Most interesting distribution of *waalaa* in Bhojpuri is its uses as nominal modifiers which in Hindi are generally manifested by relative clause construction. We have listed only some of the most common uses of *waalaa* in Bhojpuri because of a paucity of written data for verification. More detailed work is needed to list all the distributional behavior of *waalaa* and to do its functional categorization in Bhojpuri. There is scarcity of written corpus for Bhojpuri and the data used for this work have been collected from different speakers of the language. We also observe that for a detailed analysis of *waalaa* in Bhojpuri, we need data from different varieties of Bhojpuri.

References

Bhatia, T. 2008. *Colloquial Hindi: The Complete Course for Beginners*: Routledge

Shapiro, M.1989. *The Primer of Standard Hindi*. Motilal Banarasidas. Delhi.

Sharma, . 1958. *A Basic Grammar of Modern Hindi*. CHD, Govt. of India.

Case marking of Asamiya in comparison with Bangla for animate and inanimate objects

Bornini Lahiri

Centre for Linguistics, Jawaharlal Nehru University
New Delhi, India
lahiri.bornini@gmail.com

Abstract

The present paper focuses on case marking in animate and inanimate objects of Asamiya in comparison with Bangla. These two eastern Indo-Aryan languages share many features. If one looks at the case markings of the two languages then it can be seen that they share most of the cases though they are realized differently through affixation and postpositions. Both, Asamiya and Bangla perceive differences between animate and inanimate objects through various case markings. Though both the languages follow a similar pattern to differentiate between animate and inanimate objects through case marking yet there lie differences like Asamiya uses two different markers for animate and inanimate objects in instrumental case while Bangla uses the same with an additional accusative marking for the animates. In this paper I have tried to study various case markings of Asamiya and Bangla distinguishes animates and inanimates and I have also looked upon the instances where a particular case marking differentiates between the two, in one language but not in the other.

1 Introduction

Asamiya and Bangla, the two of the eastern Indo-Aryan languages are quite similar to each other. In fact till recently both were taken as one language by many linguists. They are the members of the eastern Indo-Aryan language and are descendent of Magadhi Prakrit. Along with other similarities they share same case features, though they are realized differently through affixation and postpositions. They share usage of non-cannonical case marking as well as the feature of differentiating animate and inanimate object through case marking.

Cases are traditionally described as "*a system of marking dependent nouns for the type of relationship they bear to their heads*" (Blake 1994). Generally case is marked by an affix or a complete word. At times both are needed to realize one case;

e.g: bristi-r   jonne   (Bangla)
           rain-GEN   reason   (word-to-word)
Due to rain   (free translation)

In Bangla /jonne/, the postposition for causative cannot come without genitive marker /-r/. Together they give the meaning of causation. In some instances, one case-marking may serve different cases at different times, like in the above example genitive marking along with the postposition is used to give causative meaning but is not used to show possession.

The primary function of case marking, be it via affix or adposition, is typically seen to be a relational one, namely of denoting the nature of the semantic dependency. The more unpredictable the semantic nature of the dependency relation, the more likely it is to be overtly marked by case. In addition to this primary relational function case marking may also perform two other functions. The first of these is the discriminatory or differentiating function (Comrie 1989). It is the use of case to distinguish the A (agent) from the P (Patient) in transitive clauses and the R (recipient) from the T

(theme) in ditransitive ones. The second additional function that case marking may fulfill is that of the indexing properties of the referents of arguments or of the clause itself. The properties of the referents of arguments may concern their inherent characteristics (e.g. animacy) or contingent ones (e.g. definiteness, referentiality, focus).

There are basically two approaches for case marking feature; the **discriminatory** approach and the **indexing** approach. According to the discriminatory approach "*the basic function of core case marking is to overtly distinguish between the arguments of a transitive clause*" (Naess 2007). It says that generally overt case marking is used in two argument clauses to distinguish between subject and object like ergative case is used on subject and accusative case on object. Whereas indexing, approaches case marking acts as the reflector or the index marker of certain semantic properties like animacy, humanness or definiteness.

There are many languages across the world which differentiates objects on the basis of factors like definiteness, animacy and specificity. Hebrew uses overt case marker on definite objects, but not on indefinite ones; while Catalan case marks human objects (pronoun objects), but not objects referring to non-human (Naess, 2007). Generally, the distinction of these factors is shown by overtly case marking one factor while not marking the parallel of it. In Asamiya and Bangla, case markers distinguish between animate and inanimate objects but not only by marking the one and leaving the other unmarked but also by marking the two (animate and inanimate) with different markings.

2 Case markers of Asamiya and Bangla

Asamiya overtly marks five cases, namely agentive /-e/, genitive /- r/, accusative /- k/, locative /-t/ and instrumental /-ere/, /-di/, /-dara/, /-re/. Bangla also case marks all these cases except the agentive case. For Bangla, the case markers are /-er/, /-der/(plural human), /-kar/, /-ker/(for deixis) for genitive, /-ke/ and /-te/ for accusative and locative respectively and /dije/ for instrumental.

Accusative case

Accusative (ACC) case marker works in a similar way in these languages. It is generally used with animate objects in both the languages.

eg1 (a)	**ram**	**hori-k**	**dekhile**	(Asamiya)
	Ram	Hari-ACC	saw	(word-to-word)
	Ram saw Hari			(free translation)
eg (b)	**ram**	**am-to**	**dekhile**	(Asamiya)
	Ram	mango-CLA	saw	(word-to-word)
	Ram saw the mango			(free translation)
eg (c)	**ram**	**am- k**	**dekhile***	(Asamiya)
	Ram	mango-ACC	saw	(word-to-word)

In Bangla also one can find parallel examples.

eg2 (a)	**ram**	**hori-ke**	**dekhlo**	(Bangla)
	Ram	Hari-ACC	saw	(word-to-word)
	Ram saw Hari			(free translation)
eg (b)	**ram**	**am-ta**	**dekhlo**	(Bangla)
	Ram	mango-CLA	saw	(word-to-word)
	Ram saw the mango			(free translation)
eg (c)	**ram**	**am-ke**	**dekhlo***	(Bangla)
	Ram	mango-ke	saw	(word-to-word)

From the examples of both the languages it can be seen that when the object is inanimate then the case marker is dropped. In some other examples also we can witness this feature.

eg3 (a)   **l ra – tu – k   tebil- r   upor- t   thoa**   (Asamiya)
                boy-CLA-ACC table-GEN on-LOC   keep   (word-to-word)
                Keep the baby on the table.   (free translation)

eg (b)   **kitab-khon   tebil-r   upor- t   thoa**   (Asamiya)
                book –CLA   table-GEN on-LOC   keep   (word-to-word)
                Keep the book on the table   (free translation)

eg (c)   **kitab-khon- k   tebil-r   upor- t   thoa#**   (Asamiya)
                book –CLA-ACC table-GEN on-LOC   keep   (word-to-word)
                Keep the book on the table   (free translation)

eg4 (a)   **bacca – ta - ke tebil-er   opor-e   rakho**   (Bangla)
                child-CLA-ACC table-GEN on-LOC rakho   (word-to-word)
                Keep the baby on the table.   (free translation)

eg (b)   **boi-ta   tebil-er   opor-e   rakho**   (Bangla)
                book –CLA   table-GEN on-LOC   keep   (word-to-word)
                Keep the book on the table.   (free translation)

eg   (c)   **boi-ta-ke   tebil-er opor-e   rakho#**   (Bangla)
                book –CLA-ACC table-GEN on-LOC keep   (word-to-word)
                Keep the book on the table.   (free translation)

Example 1c and 2c are ungrammatical whereas example 3c and 4c are not ungrammatical but they are not good sentences of Asamiya and Bangla respectively. This is so because when an inanimate object takes classifier and then takes accusative case marking then the sentence does not become ungrammatical in these languages though the sentence also does not remain a standard sentence. But when accusative marking is attached to an inanimate object without attaching classifier to the object then the sentence becomes ungrammatical.

At times, in both the languages accusative marking is used with an inanimate object with a classifier. This type of structure is generally used to give animate reading of an inanimate object (e.g. to show degradation of an animate object or used with the child).

eg5 (a)   **ram   bo tu-ti-ke   dekh-lo**   (Bangla)
                       Ram   thing-CLAS-ACC see-1st PER SING (word-to-word)
                       Ram saw the thing (animate).   (free translation)

Here the word "thing" (boStu) is used for some person but to make fun of him/her or to insult him/her. In the following example one can see that accusative marking is being used with an inanimate object, doll. The doll may be animate for a child though it is inanimate for the rest of the world.

eg6 mita   **putula-zoni- k   dekhile.**   (Asamiya)
               mita   doll-CLA-ACC   saw   (word-to-word)
               Mita saw the doll   (free translation)

Here it should be noted that 'the doll' is taking the classifier /zoni/ which is used for female animates. Generally 'the doll' takes the classifier /tu/ but here it is referred as an animate object so along with the classifier for female(animate) it also takes the accusative marking.

eg7   **putul-ta-ke   dekhli?**   (Bangla)
              doll-CLA-ACC   see?   (word-to-word)
              Did you see the doll?   (free translation)

Locative case

The distinction between animate and inanimate is also maintained through locative (LOC) case marking. Asamiya uses /- t/ and Bangla uses /-te/ for locative marking. Here the situation is just opposite to accusative case marking. Locative case marking comes with inanimate objects but not with animate objects. When it comes with animate objects then it does not give locative reading.

eg8   **kitap-khon   bakos- t   ase**   (Asamiya)
               book-CLA   box-LOC   be   (word-to-word)
               The book is in the box.   (free translation)
eg9   **boi-ta   bak s-e   ace**   (Bangla)
               book-CLA   box-LOC   be   (word-to-word)
               The book is in the box.   (free translation)

Locative marker cannot be directly attached to an animate object. To get locative meaning out of an animate object locative marker is attached to certain locations like

        majh- t (Asamiya)   majh-e (Bangla)   (middle+LOC=between),
        upor- t (Asamiya)   opor-e (Bangla)   (top+LOC =on top)

These locations come after genitive markers to give locative meaning for the animate objects. In Asamiya without the location also the locative marker can be used. But then the genitive marker should precede locative marker. Directly, locative marker cannot be attached to the animate object.

eg10 (a)   **tumi mu-r-tat   ki   dekhile?**   (Asamiya)
        you   I-GEN-LOC what see   (word-to-word)
           What you saw in me?   (free translation)

eg10 (b)   **tumi   mu- t   ki   dekhile?***   (Asamiya)
         you   I-LOC   what see   (word-to-word)
eg11 (a)   **tumi ama-r modh-e ki   dekhle**   (Bangla)
        you   I-GEN in-LOC what see   (word-to-word)
          what you saw in me?   (free translation)
eg11 (b)   **tumi ama-e ki   dekhle***   (Bangla)
         you   I- LOC what see   (word-to-word)

In Bangla if locative marker is used with animates, it gives different meanings but not that of locative. Some of such meanings are that of generic (eg11(a)), coupled reciprocal (eg11(b)) and dative-accusative (eg11(c)).

eg12 (a):   **kukur-e   kamre dae.**   (Bangla)
         Dog-LOC   bite   give   (word-to-word)
           Dogs bite   (free translation)
eg12 (b):   **dui bondhu-te   jhogra**   (Bangla)
         two   friend-LOC   quarrel   (word-to-word)
         There a dispute between the two friends.   (free translation)
eg12 (c):   **o   ama-e/-ke   bollo**   (Bangla)
         he/she I-LOC/-ACC said   (word-to-word)
           He/she told me   (free translation)

Instrumental case

In instrumental (INST) case marking the differentiation between animate and inanimate is shown by two different case markers in Asamiya. It is marked in Asamiya by the suffix /-ere/ & /-di/ when the object is inanimate & /-dara/ when the object is animate.

eg13:   **kamkora-zuali-joni-dara   sapha kora**   (Asamiya)
          working-girl-CLA-INST   clean   do   (word-to-word)
          Get it cleaned by the maid.   (free translation)
eg14:   **jaru –ri   sapha kora**   (Asamiya)
          broom-INST   clean   do   (word-to-word)
          Get it cleaned by a broom   (free translation)

In Bangla when the object is inanimate then the postposition /dije/ is used. But when the object is animate then accusative marking /-ke/ is used before /dije/.

eg15:   **jharu dije   jhat   deao**   (Bangla)
          broom INST   sweep give   (word-to-word)
          Get it swept by the broom.   (free translation)

eg16:   **dai-ke   dije   jhat   deao**   (Bangla)
          maid-ACC   INST   sweep give   (word-to-word)
          Get it swept by the maidservant.   (free translation)

3 Conclusion

In Asamiya out of the five case markers three cases distinguish between animate and inanimate objects. For accusative case it drops its case marking when the object is inanimate while in locative marking, it is dropped when the object is animate. In instrumental case two different case markings are used for animate and inanimate objects.

In Bangla also the situation is quite similar. Like Asamiya it also drops its accusative marking when the object is inanimate and in locative marking the marker is dropped when the object is animate. In Bangla case markers of every case except genitive case distinguishes between animate and inanimate objects. Bangla genitive case marker distinguishes between plural human and non-human. For plural human /-der/ is used while for all other instances /-r/ is used.

eg17 (a)   ama-der   bari   (Bangla)
             3PER-GEN   house   (word-to-word)
                Our house.   (free translation)

eg17 (b)   kukur- gulo-r   bari   (Bangla)
             dog- CLA-GEN   house   (word-to-word)
                Dogs' house   (free translation)

In the following table the case markers which distinguish between animate and inanimate in Asamiya and Bangla, are shown. In the chart it is clearly visible that the two languages in a similar way distinguish between animate and inanimate objects.

	Asamiya		**Bangla**	
Case marking	*Animate*	*Inanimate*	*Animate*	*Inanimate*
Accusative	-ɔk	Not used	-ke	Not used
Locative	Location + LOC or GEN+LOC	-ɔt	GEN+Location +LOC	-te
Instrumental	-ere	-de	-ke + diye (PP)	dije

Table: 1 showing case markers of Asamiya and Bangla

In many languages object is marked for specific semantic features. The concept of Differential object marking (DOM) is also similar which says that there is a wide range of languages where object case marking differs according to certain properties of object (Naess 2007). It can be seen that this characteristics is also found is Asamiya and Bangla.

It can also be said that the cases show perception of the language so the same relation between two objects in different languages is realized through different cases. It can be said that Asamiya and Bangla perceives difference between animate and inanimate objects which along with classifiers is also realized through cases.

References

Abbi. Anvita. 1994. *Semantic universals in Indian languages*. Shimla: Indian Institute of Advanced Study

Blake. B. J. 1994.*Case.* Cambridge: Cambridge University Press.

Comrie. Bernard. 1989. *Language universals and linguistic typology*. Oxford:Blackwell.

Dasgupta. Probal. 2003. 'Bengali'. *The Indo-Aryan languages.* edi: Cardona G & Jain D

London: Routledge Publication

Goswami. S. 2003. 'Asamiya', *The Indo-Aryan languages*. edi: Cardona G & Jain D

London: Routledge Publication

Masica. C. P. 1991.*The Indo-Aryan Languages*. Cambridge: Cambridge University Press.

Masica. C. P. 1976. *Defining a Linguistic Area*: *South Asia*. Chicago: University of Chicago Press

Naess. Ashild. 2007. *Prototypical Transitivity.* Amsterdam: John Benjamins Publishing Company.

Implementation of Transfer Grammar in Telugu-Hindi Machine Translation

Christopher Mala
Center for Applied Linguistics and Translation Studies
University of Hyderabad
LTRC,IIIT-Hydearbad, Gachibowli
chirstopher.mpg08@research.iiit.ac.in

Abstract

This paper describes the experiments on Transformation of Grammar from one language to another while translating text through machine. It is known that every language has its own phenomena and its own way of representation. While translating from one language to another it is very important to retrieve these language phenomena information of target language from source language, which might not be there in the source language. These language dependent phenomena can be seen when we are translating languages of cross language families. In this paper we have tried to explain how grammar is transferred from Telugu (Dravidian language family) to Hindi (Indo-Aryan family).

1 Introduction

1.1 Transformational Grammar (TG) Definition

Transformational grammar seeks to identify rules (transformation) that govern relations between Chunks of a sentence, on the assumption that there exists a fundamental structure beneath the word order. Transformational grammar is the starting point for the tremendous growth in linguistic studies since 1950s.

1.2 Why Transformation Grammar is Required

The usual usage of the term 'transformation' in linguistics refers to a rule. For example, a typical transformation in TG is the operation of subject-auxiliary inversion (SAI). This rule takes as its input a declarative sentence with an auxiliary: "John has eaten all the heirloom tomatoes", and transforms it into "Has John eaten all the heirloom tomatoes?". These rules were stated as rules that held over strings of either terminals or constituent symbols or both X NP AUX Y   X AUX NP Y (where NP = Noun Phrase and AUX = Auxiliary). Transformations are no longer structure changing operations at all, instead they add information to already existing trees by moving constituents. The earliest conceptions of transformations were that they were construction-specific devices. A different transformation raised embedded subjects into main clause subject position in sentences and yet a third reordered arguments in the dative alternation. With the shift from rules to principles and constraints, these construction specific transformations are morphed into general rules. Generalized Transformations (GTs) take small structures which are either atomic or generated by other rules, and combine them.

1.3 Rules and Description

A formal Linguistic operation which enables two levels of structural representation, Dependency parsing and Phrase Structure, which contains sequence of terminals and non-terminals. Where as a Transformational Rule consisting of a sequence of symbols rewritten, as equivalent corresponding sequence to the source language. The input to Rule is the Structural Description, which defines the class of Phrase-Markers to which the rules can apply. The rule then operates a Structural Change on this input, by performing operations that were instructed in the rule. Transformation (Movement)

modifies an input structure by reordering the elements it contains. When this operation is seen as one of the moving elements to adjoin positions in a phrase-marker, it is known as Adjunction. Insertion(Transformation) add new structure elements to the input sentence. Where as Deletion(Transformation) eliminates elements from the input sentence. Several models of transformation grammar have been presented since its first outline, that can manage some of the below listed functions. a) Syntactic components b) Phonological Components c) Semantic components. To design these grammar rule, we need to have strong knowledge about the source and the target languages. It is very important to understand the divergence between the two languages. Divergence at various levels like Lexical level, Morphological level and Syntactical level. Transformation Grammar(TG) deals with both Morphological and Syntactical divergence. TG is necessary in Translation to resolve the divergence between languages and produce translated text which is syntactically and semantically correct. Here we formulate few rules for the language that are of two different families. Taking into consideration of the structural and semantic divergence of the both languages, it has been tried to formulate transfer rules for different sentence from Telugu to Hindi. In this we build rules by hypothesizing and then generalizing over them. These generalized rules represent contexts with constraints over semantic categories. We need to classify language divergence into various categories in different terms, all these divergence can be resolved by a set of TG rules. We can classify TG rules into Major and Minor. Some of them are:

- Copula
- Ergative
- Participles ("yA huA", "nA vAlA")
- Conjuction (Ora)
- Modifying verb into Finite Verb
- ani/ Complementizer
- Disjunction elements
- Discourse Markers

These are grouped into four and are explained :

- Adding of Copula and other language specific data.
- Deletion of Grammar that is not required in the target language.
- Modification of the source language Grammar according to target language .
- Smoothing of the target language Grammar.

In this paper it has also been explained that Transfer Grammar engine which is of language independent and it can be used by training with rules. This study is being used in Indian Language - Indian Language Machine Translation project (IL-ILMT system) which is funded by Govt. of India (Minstery of Information Technology) being developed at CALTS lab in University of Hyderabad under the guidance of Prof. G. Uma Masheshwar Rao, Head, CALTS, HCU.

2 Introduction to Languages and their divergences

Telugu as a language belongs to South-Central group (SD-II) of Dravidian languages. Morphologically Telugu is agglutinating in structure. There are no prefixes or infixes. Grammatical relations are expressed only by suffixation and compounding. Syntactically all Indian languages are of OV type, head-right-final and right-branching. The subject argument is generally expressed by a noun phrase (NP), but a post-position or case phrase with the head nominal in the dative case can also function as the subject, latter called as 'dative subject sentence'. The predicate has either a verb or a nominal as head. Sentence with nominal predicate is equivalent sentence, which lack the copula or the verb 'to be' in Telugu. Nominal and verb predicates have different negative words which express sentence negation. A negation word is an inflected verb meaning 'to be' or 'to be not'. But

this cannot be seen in Hindi, we can see the negative words as separate lexical items. Non-finite verbs, which head sub-ordinate clause, have affirmative and negative counter parts in Telugu . The arguments of NPs which occur as complements to a verb, are derive from the semantic structure of a verb; for instance, an intransitive verb require only one argument Agent/Object, where as transitive verb requires Agent+Object: a causative verb requires, Agent(causer) + Agent(casuse)+Instrument+Object. The passive voice is rarely used in modern Dravidian Languages.

Source Side Analysis (SL)
⇓
Transfer Grammar
(SL-TL)
Lexical Substitution
⇓
Target Side Generation (TL)
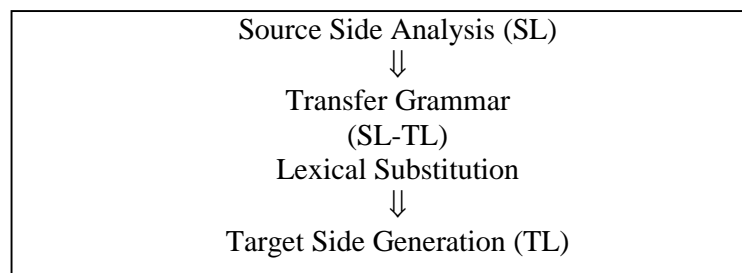
Table 1: Structure of MT.

3 How to use T.G in MT

3.1 Flow of MT
After analysing the input text in the source side. It has to be passed for lexical transfar. Before passing to lexical transfar, the process of transfar grammar should be done. As shown in the above fig.

3.2 Transfer Grammar Rule Format Specifications
A grammar is a way to formally describe the structures of a language through a set of rules. Several formalisms have been developed for such descriptions in the field of NLP. PSG is a purely syntactic approach which uses a set of phrase structure rules to write the grammar of a language. It is constituency based and the order of elements in a sentence is implicit in it. DG, on the other hand, tries to capture the semantic relations of the elements in a sentence. For writing the transfer grammar rules a rule format needs to be specified. And since Indian languages are structurally very similar it is possible to achieve a high degree of correct transference without going to a deeper level of sentence analysis, i.e. a fully parsed sentence. Therefore, the transfer grammar format should also be able to handle shallow parsed inputs. For this level, the TG have rules that take chunks (for PSG) or bags (for DG) as inputs. For some special cases, a simple parsed (see below) level can also be accepted. The rules would be stated differently in the PSG and DS formalisms. Conventions need to be defined for both these formalisms. However, before going into specifications of rules in a particular format it is important to identify the rule requirements. The transfer grammar rules would be stating the structural changes from the (Source Language) SL to (Target Language) TL. Rules would have an LHS and an RHS. The format of a transfer grammar rule would have two parts the Left Hand Side (LHS) part and the Right Hand Side (RHS) part. Therefore, the format of the rule is LHS => RHS A Left Hand Side (LHS) and a Right Hand Side (RHS) which are separated by the symbol '=>'. The symbol '=>' stands for 'transfer to'. The LHS has the input from the source language Telugu in this case and the RHS has the expected output of the rule for the target language. Therefore, the rule states that if the source language has a structure with two NPs in a sequence and they are related to each other by a genitive relation then a genitive marker should be inserted on the RHS. This is stated by changing the value of the attribute 'cm' from LHS (cm-

UNDEF) to RHS (cm=kI). Ex: NP 1(({<case=gen, cm=UNDEF>})) NP2=> NP 1(({<case=gen,cm="kI">})) NP2

4 Adding of target language specific data (Copula and ergator)

In this, data has handled, that is missing in the source language but it is very necessary in the target language to get proper translation. A few of the things are discussed below.

4.1 Handling of Obligatory Transformation
As it is known that the oblique form for common nouns in Telugu take "*ti*" as case maker (*oVMti*, *iMti*) for proper nouns its oblique form is *du* (*rAmudu*). But in Hindi there is only one case marker for oblique nouns (*kA*).
   Rule:
      NP 1(({<case=o,tam=ti>})) NP 2 =>
      NP   1(({<case=o,cm=kA>}))   NP2

4.2 hE insertion
Noun phrase (NP 1) is followed with an Adjective (NP 2) in source language (SL Telugu), but in Hindi we need a copula in the target language at the end of the sentence.
Ex: (Tel) *rAmudu*   *maMcivAdu*.
        (HIN) *rAma   accA   vAlA   hE*.

The rule for the above example is given below:
Rule:
   NP1NP2(({<lcat=adj>}))=>NP1NP2+VGF(({hE%VM<root=hE,lcat=v,gen=m,num=sg,per=hE,tam=hE>}))

4.3 Example 2
If there is no verb in the source side then insert hE before the sentence ends.
Ex: *(Tel) rAma lakRamaNulu annaxammulu.*
        (Hin) rAma lakRamana BahI hE.
Rule:
      VGF(({<.%SYM>}))=>VGF((+{<root=hE,lcat=v,tam=hE>}{%SYM}))

4.4 Ora insertion
If there are two noun phrases (NP 1, NP 2) with any long vowel as case marker then a conjunction is inserted in between these two noun phrases.
Ex: *(Tel)   I waragawilo kurchIlu ballalu unnAyi.*
        (Hin)   yaha kakRa me kursi Ora meja hE.

   Rule:
NP1(({<cm="A">}))NP2(({<cm="A">}))=>NP1(({cm="0"}))+CCP(({OraCC<root="Ora",lcat="conj">}))NP2(({<cm="A">}))

4.5 ne insertion
A direct noun phrase (NP1) is followed with an oblique noun phrase (NP2) and a verb phrase (VP) in the source side. Then in the target side ne is inserted in the first noun phrase (NP1). And rest are retained.

Ex: *(Tel)   rAmu puswakaM caxivAdu.*
         (Hin)   rAma ne puswaka paDA.

Rule
NP1((\{<case="d",cm="">\}))NP2((\{<case="o">\}))VGF((\{<tam="A">\}))=>NP1((\{<case="o",cm="ne">\}))NP2((\{<case="o">\}))VGF((\{<tam="A">\}))VGF((\{<tam="A">\}))

5 Deletion of Grammar that is not required in Target side

In this we are trying to frame rules to delete theinformation that is required in the target language from source language. Here are some of the examples that explain how deletion is done in Transformation Grammar.

5.1 Example:1
The word *samayAnni* in telugu will be having root as *samayaM* and case marker as *ni*, but where as in hindi it is *samaya* with case marker as 0. So the case marker is dropped in target side. Case marker *ni*, and word ending with $x.aM, and lexical category *noun*, can be dropped in target side.
Rule:
NP((\{<root="samayaM",cm="ni">\}))=>NP((\{<root="samayaM",cm="0">\}))

5.2 Example:2
If a 3rd person pronoun is having case marker as *ki* in source side, then it should be dropped in the target side. Rule: NP((\{<root="ixi",lcat="pn",cm="ki">\}\{gAnu%RP\})) => NP((\{<root="isa",cm="0">\}\{gAnu%RP\}))

6 Modifying in the Source side Grammar according to Target side

In Telugu, any finite verb is ending with *-ani*, example a verb like *ceVppamani (keha kara)* where *ceVppu* is the base form. The participle *-ani* means *kara*. In Telugu we can see this *-ani* within the word, but Hindi *kara* is and aux-verb. So it has to be denoted as post position to the main verb (VM).
Rule:
VGF((\{<tam="$x-ani">\}))=>VGF((\{<tam="$x">\}))+NP((\{aniPSP<root="ani",lcat="psp">\}))

6.1 Example 2:
In this example we can see a direction nominal which case marker is as *na*(nominative), this case marker is converted in locative marker *lo* in the VGF((\{<tam="A">\}))

6.2 Example 3:
If any non-finite reduplicated verb is occurred in the sentence, eg: *ceVppi ceVppi/cUsi cUsi/wini wini*, we even have the tense reduplication also. But in Hindi, we can see the tense reduplication is not possible. And the appropirate word for this reduplication verb is *bawA bawA kara*. So here one of the tense marker is dropped in the source side.
Rule:
VGNF1((\{<lcat="v",tam="i">\}))VGNF2((\{<lcat="v",tam="i">\}))=>VGNF1((\{<tam="0">\}))VGNF2((\{<tam="i">\}))

7 Smoothing of the target language Grammar

In Telugu, verb phrase like *caMpina puli, winina palYleVM* etc. have a lot of ambiguity. Lets take example *caMpina puli* which mean *mAra ne vAlA Sera*", have tam as *-ina* for the verb main in the verb group. It is know that for the tam *-ina*, its corresponding hindi has two values, *hE-jo-vaha*, *wA-hE-jo-vaha* and *yA-hE-jo-vaha*. Depending on the context we have to choose the correct tam. When we have *yA-hE-jo-vaha* we should compress it to *yA*, and when it is *wA-hE-jo-vaha* it is substuted with *ne-vAlA* Example 1 VGF(({<lcat="v",tam="wA-hE-jo-vaha">})) => VGF(({<tam="ne-vAlA">})) Example 2 VGF(({<lcat="v",tam="yA-hE-jo-vaha">})) => VGF(({<tam="yA">}))

8 Conclusion and Evaluation

8.1 Evaluation

A test a made to know the Understandability of the text after implementaing the T.G in the IL-ILMT system. For this test we have choosen 30 sentences in different construction. They were executed through MT system and given for manual evalution. Their response is given below.

Reader -1	Rank 1-2	Rank 3-5	Accuracy
Before (TG)	14	16	53.33
After (TG)	19	21	70.00

Table 2: Data-Sheet of Reader-1.

8.2 Conclusion

As we know that every language has its own phenomena which are called as language divergences. These language divergences should be taken care of while translating the text form one language to another. This should be even more carefully handled when the translation is across language families.

These divergences can be at the **Lexical level, Morphological level and Syntactic level**. Transformation Grammar (TG) deals with both Morphological and Syntactic divergences.

TG vitally reduces these divergences while translating text and increases the understandibility of the reader.

References

Krishnamurti Bh. 2003. *The Dravidian Languages*. Cambridge University Press, Cambridge, UK.

Krishnamurti Bh and Gwynn J.P.L. 2003. *A Grammar of Modern Telugu*. Oxford University Press, Delhi, India.

Robert Coldwell. 1891. *Comparative Grammar*.

Beames, John. 1966. *A Comparative Grammar of Modern Aryan Languages of India*. Delhi.

Gildea, Spike. 2000. *Reconstructing Grammar: comparative, linguistics and grammaticalization*. John Benjamins Publishing Company.

Sjoberg, Andree F. 1992. *The Impact of the Dravidian on Indo-Aryan: An Overview*.

Masica, Colin Paul. 1996. *The Indo-Aryan Languages*. Cambridge University Press Cambridge,UK.

Automatic Extraction and Incorporation of Purpose Data into PurposeNet

P. Kiran Mayee   Rajeev Sangal   Soma Paul
kiranmayee@research.iiit.ac.in   sangal@iiit.ac.in   soma@iiit.ac.in

Language Technologies Research Centre
International Institute of Information Technology
Hyderabad, India

Abstract

PurposeNet is a knowledge base of objects and actions in which the knowledge is organized around purpose. Such knowledge also connects with language – namely, verbs for related actions. It can be used with an embedded reasoner, resulting in an effective system for QA, topic-listing, summarization and other tasks. However, extracting PurposeNet related data manually is time-consuming, labor-intensive, and expensive. This paper describes a framework for automatic purpose data extraction, given a corpus. It identifies a set of lexico-syntactic patterns that are easily recognizable, that occur frequently and across text genre boundaries, and that indisputably indicate the lexical relation of purpose data. It also deals with the subsequent automatic incorporation of this data into the PurposeNet resource. The results are used to augment and critique the structure of a large hand-built resource. The extent of success, in terms of richness of the resource, achieved in the process is also discussed.

1 Introduction

Today, an increasing need is being felt to develop technology to help us manage and make sense of the information overload. One method to do so is the efficient use of commonsense knowledge. It is assumed that every person possesses commonsense. Such knowledge is typically omitted in communications, such as text. NLP requires a surprisingly large amount of commonsense, which currently only people possess. It is our purpose to find ways to provide such commonsense to machines. The methodology adopted is to create a network of artifacts related through purpose, creating a closure, which, then, could be used in conjunction with a state of the art reasoner, thereby bridging the manmachine gap and providing commonsense knowledge to NLP.

1.1 Purpose

Purpose, taadarthya or aim in its most general sense is the anticipated result which guides action. For example, when an umbrella is opened, it is anticipated that it will protect someone from the elements, i.e., the anticipated result of protection guides the action of opening. Purpose provides a ready principle to organize knowledge. It is assumed that "Every artifact has one and only one primary purpose".

1.2 PurposeNet

PurposeNet is a knowledgebase dealing with specialized attributes of artifact – namely, its purpose, purpose of its types, its components, and accessories, and their purposes, in turn, as also data about its birth, normal processes, sideeffects, maintenance and result on destruction. For example, given an artifact called vehicle, its purpose is to transport people or goods. Vehicles could be of airborne, waterborne or landborne type. Each of these again has their purposes. Land vehicles have components such as wheels, engine, body, brakes etc. Each of these components is again an

artifact. Vehicles are born or manufactured by assembling the components. A vehicle can start, move, pause and stop. These, then, are its processes. Vehicle parts get worn out, which constitute sideeffects. Finally, different results are obtained with respect to its leftovers when a vehicle is discarded. The requisite database for the PurposeNet has been constructed in a way that is understandable by humans and processable by computers.
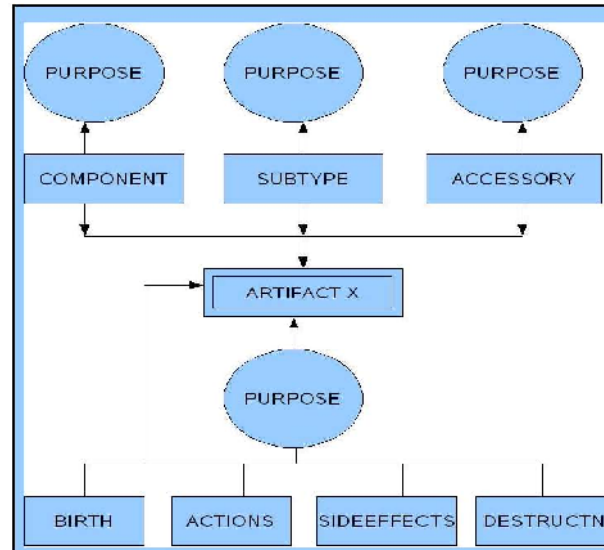

Fig 1. PurposeNet showing purpose for artifact X, the corresponding artifact needed to implement this purpose, subtypes, components and accessories of the artifact. It also shows the purposes of its subtypes, components and accessories, as also all processes with respect to to its purpose. (The squares represent artifacts. The broken squares represent processes – birth, actions, side effects and destruction, and circles represent purpose. Arrows connect the different processes to their purpose.)

1.3 Building the PurposeNet
 PurposeNet building involved a sequence of six phases. Firstly, the appropriate template was designed for representing each of the artifacts and their corresponding hierarchies. This was followed by revising and refining of the template, so as to avoid redundancies. Phase three involved selection of appropriate domains and information retrieval from the web and other sources. After this, we proceeded to create the corresponding web ontology. Finally, the created ontology was tested using a reasoner. While doing so, the following major issues were taken care of: adding a new domain in the existing ontology network, changing the knowledge format, performance issues and guaranteed quality of service, scalability and ease of making any modifications.

1.4 Need for Automation
   The manual extraction of data and building of PurposeNet ontology is quite exhaustive as several experts are required to put in hours of browsing to find the data corresponding to the concerned attribute and to incorporate it – also called the knowledge acquisition bottleneck [Ratsh et. all, 2003]. This also leads to a slow progress in the creation of a knowledge base that may be as big as a million artifacts. Text documents are, and will always be massively available. Hence, a methodology was designed and executed for the automatic extraction of this massive data and the subsequent ontology population of the extracted data.

1.5 Principle

Languages have special cues to recognize purpose data. Thus, if apriori information about the cues is known, and if it may be pointed out beyond doubt that the sentence pertains to an artifact, then, the purpose of the artifact may be easily determined, provided it exists in the given sentence. Purpose knowledge is extracted and learned by automatic pattern matching of sentences of the relevant corpora. For example, the Princeton corpus sentence "Stone chisels are used to carve or cut stone" contains data about the artifact, its purpose and the entity upon which it acts. An efficient pattern matching algorithm may identify and extract all such relevant data.

Languages have special cues to recognize purpose data. Thus, if apriori information about the cues are known, and if it may be pointed out beyond doubt that the sentence pertains to an artifact, then, the purpose of the artifact may be easily determined, provided it exists in the given sentence. Purpose knowledge is extracted and learned by automatic pattern matching of sentences of the relevant corpora. For example, the Princeton corpus sentence "Stone chisels are used to carve or cut stone" contains data about the artifact, its purpose and the entity upon which it acts. An efficient pattern matching algorithm may identify and extract all such relevant data.

2 Purpose Cues

Sentences contain clues with respect to certain semantic meaning. Usually, a word, a set of words or a set of lexical entities in a particular order tend to point to the existence of certain types of data in sentences. The clues used to determine purpose data in sentences are the purpose cues.

2.1 Classification of Purpose Cues

In building a PurposeNet, our primary goal is to extract data corresponding to – artifact name, what is its primary purpose or use, and finally, what entity it acts on. These are named respectively as <artifact>, <use> and <action_upon>. Their location in a sentence is determined by purpose cues.

A 'purpose cue' is any distinguishing pattern in a sentence that can clearly point to the existence of artifact purpose data. Several such cues may be identified. The sentences containing these cues may be broadly classified into three types – sentences commencing with artifact name, sentences containing artifact data, and sentences ending with artifact name.

Apart from these three, there are sentences that do not have any specific 'purpose cues', inspite of containing purpose data. These, then, are the exceptional cases. In the following discussion, we assume that the sentence in question contains data about the artifact and its purpose, and possibly, the template on which it acts. We may therefore categorize sentences according to their purpose cues as follows explained in sections 2.1.1, 2.1.2 and 2.1.3.

2.1.1 Sentences commencing with artifact name

These are sentences with the cues is used to, is used for, helps in, helps to, etc. Typical examples in this category are "A brush is used to paint walls", "A sharpener is used for sharpening pencils". It may be observed that the artifact names (brush, sharpener) are matched to the beginning of the sentence. One such example is given in Fig.1, where a onetoone patternmatching is shown between the cue and an example pattern. The examples for such sentences are –
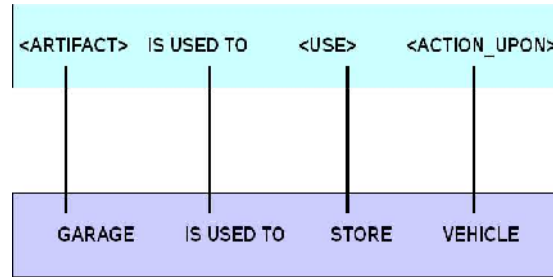
Figure 1.Illustration of Sentences commencing with artifact name.

1 <artifact> <is used to > <action> <upon> Eg : A brush is used to paint walls.
2 <artifact> <is used for> <action> <upon> Eg : A sharpener is used for sharpening pencils.
3 <artifact> <helps in> <action> <upon> Eg : Car helps in travel.
4 <artifact> <helps to> <action> <upon> Eg : Paper+Clip helps to bind papers.
5 <artifact> <is used in><action> <upon>.Eg:Oil is used in cooking.

2.1.2 Sentences ending with artifact name

Consider the sentence "Weeding is done using Scissors". These sentences end with artifact name (Scissors) and have the cue 'is done using' or 'is the use of' or 'with the help of'. Some more examples of such type are –

1 <action> <is the use of > <artifact> Eg :Fighting is the use of sword.
2 <verb> <is done using> <artifact> Eg:Weeding is done using scissors.
3 <verb> <is done with/with the help of> <artifact> Eg: Filtering is done with filter.
4 The <upon> is <action>ed with <article> <artifact> Eg:The book is covered with a cover.
5 The <upon> is <action>ed with the help of <article> <artifact> Eg:The tree is cut with the help of an axe.
6 <action> <article> <upon> <with/using> <article> <artifact> Eg:Paste the picture with gum.
7 <action> <upon> <is achieved with> <article> <artifact> Eg:Cleaning the room is achieved with a vacuum + cleaner.
8 To <action> <article> <upon>, <noun/pronoun> <need/ require/use> <article> <artifact> Eg: To paint the wall, we need a brush.

2.1.3 Sentences containing artifact name anywhere

These are sentences with the artifact name somewhere in between the words of the sentence. For example "Use the scale to draw a line" and "Utilize the airpump to fill the tyre". It may be observed that the artifact names (scale, air + pump) are embedded within the sentence/pattern. Cues such as Use < artifact> to <use> <action_upon>, use of <artifact> is to <use> <action_upon>, Utilise < artifact> to, etc. are used. The format of sentences and examples where artifact names could appear anywhere are.

1 The <purpose/ use/utility> of <artifact> is to <action> <upon>.
   Eg: The purpose of a bag is to carry things.
2 The <use of> <artifact> <is to> <action> <upon>
   Eg: The use of a bucket is to carry liquid.
3<Use> <article> <artifact> <to> <verb> <article> <upon>
   Eg: Use the scale to draw a line.
4<Utilise> <article><artifact> <to> <action> <article><upon>
   Eg: Utilise the air+pump to fill the tyre.
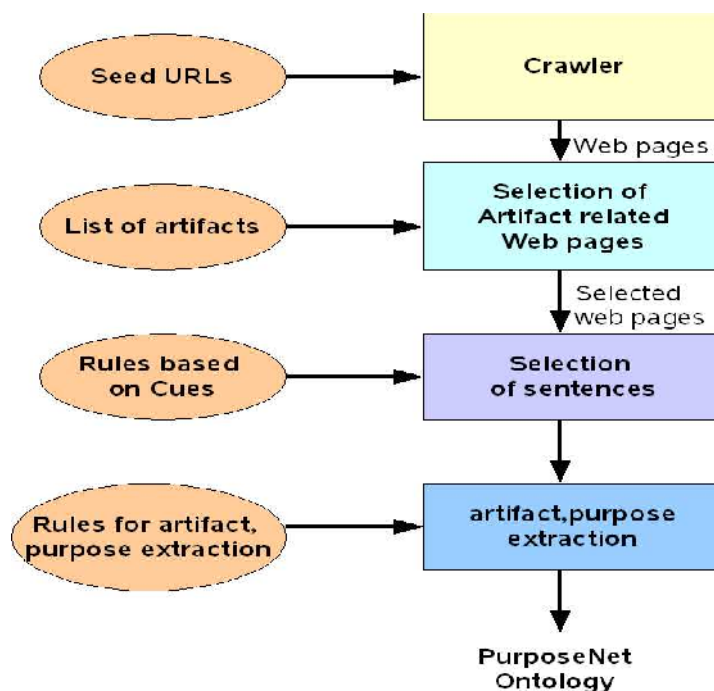
Figure 2 Methodology for Purpose data extraction and incorporation into PurposeNet

2.1.4 Exceptional Cases

There are cases such as that of a simple sentence representing the purpose of an artifact. Since there are no cues, in such cases, the purpose data cannot be extracted. such as

   1. <artifact> <action> <upon>   { In present tense}

       Eg : Pen writes on paper.

3 Methodology for Purpose Data Extraction

Rule based method is used to obtain purpose data. Rules themselves are handcrafted by looking at corpus files downloaded form the web. These files are given as input to a patternmatching algorithm, which extracts sentences containing purpose data of artifacts based on the rules which include occurrence of cues in the sentences. Data pertaining to the name of the artifact, its purpose and the entity on which it acts can now be determined by a second, more stringent pattern matching. The data so obtained can then be used to populate the Ontology. Fig.2 illustrates the the purposesentence collection module collects sentences from web/ text files. The purposeextraction module determines the artifact, action and action_upon data, passing on the same to the ontology creation module for incorporation into the corresponding ontology.

3.1 Algorithm for purpose data extraction from corpus

Given a file with 'purpose sentences', the PurpDataExtract algorithm extracts the required triplet data, if it occurs in the sentence. For every sentence in the corpus, this algorithm first verifies that the subject is an artifact, by making a call to the contains() module (Figure 3.). The contains() module does a pattern match with an artifact list available with it. The match() module is called next. This algorithm performs a pattern match of the sentence with the existing cue table. If a match occurs, then the data is mapped into three attributes – artifact, purpose and upon. These are then incorporated into the ontology by the standard ontology writing wizard built by the authors.

Algorithm PurpDataExtract(corpus)

Step 1 : Read first sentence in Corpus.

　Step 2 : Loop until endofcorpus –

　　2a: If contains(sentence, artifact) and match(sentence, cuetable)

　　　Then

　　　　Begin

　　　　　　extract(sentence, artifact)

　　　　　　extract(sentence, to_action)

　　　　　　extract(sentence, to_upon)

　　　　　　add_to_ontology(artifact, to_action, to_upon)

　　　　End

　Step **3**: Read next sentence

Figure **3. Algorithm for Purpose data extraction, given a corpu**s

4 Data

The data consisted of a downloaded nounartifact corpus from Princeton University containing 82,115 sentences. Also, Wordnet was downloaded, corpus size being 81,837 sentences. Finally data pertaining to 243 artifacts was automatically extracted by crawling the Wikipedia site using the Quadsucker web tool starting form 'tools' site. The cue table was created using data provided by a set of linguists, wellversed with English, who have recommended a total of seventeen cues in the three categories. The artefact list was created manually by referencing a web based dictionary.

5 Observations and Results

The results for the three corpora with respect to the efficiency of purpose data extraction may be summarized as follows –

Data from the Wordnet corpus contained the highest number of sentences with purpose data (1251), followed by the Princeton corpus (1023). The Wikipedia corpus was found to have the least number of purposerelated sentences (109 out of 243).This may be attributed to the presence of some other additional cues in the corpus.

Corpus Name	Corpus size	Number of purpose sentences	Purpose Data Density (%)
Wordnet	81837	1251	1.528
Princeton	82115	1023	1.245
Wikipedia	243	109	44.855

Table 1. Table showing summary results of purpose data extraction

The breakup of the various cue types have shown that the 'used to' cue has the highest occurrence in all the three corpuses (553 in Wordnet, 227 in Wikipedia and 857 in Princeton corpus. All other cues were found to occur at a significantly lower percentage.

Corpus Name	Purpose Hits	artifact name absent	action_upon absent
Wordnet	1251	nil	4
Princeton	1023	41	17
Wikipedia	109	44	3

Table 2. Table showing results of purpose data extraction misses

The results for density of purpose data showed that Wikipedia had a very high density of 44.9% (Table 1) in comparison to the other two. Similarly, the miss rate was highest in Princeton Corpus (58%) (Table 2). Wordnet had the highest precision (99.6) and Fmeasure (99.79).

Among the various cues, the class1 cues or the cues with artifact name at the beginning of the sentence were found most frequently in all three corpora, closely followed by the cues at the end of sentences.

5.1 Comparison with manually built ontology

The manually created Ontology had an artifact number of 7300, whereas, the artifact ontology created with semiautomated process extracted and created an Ontology of a similar kind in less than 30 seconds (for the entire Wordnet corpus).But, the Ontological error frequency was found to be higher than the manual one, since the artifact table had only 71103 artifact names in its list.

Corpus Name	Class1 (begin cue)	Class2 (ending cue)	Class3 (embedded cue)
Wordnet	70.19	0.01	24.7
Princeton	71.4	1.21	21.22
Wikipedia	84.2	01.6	12.21

Table 4 Table showing result breakup for the three cue classes

6 Issues

Redundancy was introduced by default into the Ontology. The primary purpose is not always defined in the corpus. Sentences with pronouns in place of artifact name e.g., "It is used to ..." had to be ignored. Also, cases of brand names appearing in place of common names – e.g., "The indigo is used to ... " (in place of "The car is used to ...") could not be handled using the abovementioned scheme. PurposeNet contains data other than strictly the 'purpose' of artifacts. This data could not be strictly filtered out with the given cues I.e., correctness and consistency cannot be guaranteed. The <action_upon> information was not always available in the extracted sentence. The current algorithm cannot handle a sentence containing a onetomany mapping or manytoone mapping of artifacts to purpose. Also, compound, complex and multiline manifestations of purpose data need a much more complex rulebased scheme of identification.

7 Future Enhancements

The corpus may be parsed and further refined to improve accuracy of results. Cues may be determined to automatically extract the data corresponding to the hidden cues case and for the other attributes of PurposeNet such as components, birth, actions and result on destruction. Methodologies have to be developed for handling the compound, complex, and, multiline cases. An

artifact list may be maintained for proper lookup and prevention of inconsistency. Accuracy of the extraction may be further improved by multipage lookup per artifact. Metapurpose data will have to be filtered out using some other cues. A better web crawler may be used for a more voluminous download of relevant data.

8 Conclusions

A PurposeNet Ontology was automatically created with data extracted from three sources – the Wordnet, the Princeton artifact corpus, and finally, the Wikipedia, using a web crawler. A total of 2383 purpose data for artifacts were extracted successfully using a robust methodology involving refinement and rulebased pattern matching algorithms, with no other aid. The same was incorporated into the PurposeNet resource. The time taken for the ontology development using this scheme was a fraction of that needed in manual download and manual extraction. The correctness and consistency of the ontology was not always guaranteed.

Acknowledgements

References

Asanee Kawtrakul and Chaiyakorn Yingsaeree, D. 2005. *A Unified Framework For Automatic Metadata Extaction* From Electronic Document.IADLC'05.

Carlos RodríguezPenagos , Heladia Salgado , Irma MartínezFlores and Julio ColladoVides Programa de Erik Boiy; Pieter Hens; Koen Deschacht. 2007. *Automatic Sentiment Analysis in Online Text*. Belgium Proceedings ELPUB2007 Conference on Electronic Publishing – Vienna, Austria.

Genómica Computacional, Centro de Ciencias Genómicas. 2007. *Automatic reconstruction of a bacterial regulatory network using Natural Language Processing*. BMC Bioinformatics 2007, 8:293doi:10.1186/147121058293.

Gregory A Grothaus , Adeel Mufti and TM Murali. 2006. *Automatic layout and visualization of biclusters. Algorithms for Molecular Biology*, 1:15doi:10.1186/17487188115.

Jiaming Zhan, Han Tong Loh1, Ying Liu, and Aixin. D.H.L. Goh et al. (Eds.).2007. *Automatic Text Summarization in Engineering Information Management*. ICADL 2007, LNCS 4822, pp. 347–350, 2007 © SpringerVerlag Berlin Heidelberg .

Karunesh Arora, Sunita Arora, Kapil Verma, S S Agrawal. *Automatic Extraction of Phonetically Rich Sentences from Large Text Corpus of Indian Languages*. INTERSPEECH2004 ICSLP Jeju, Korea.

KunWoo Yang, SoonYoung Huh. 2008. *Automatic expert identification using a text categorization technique in knowledge management systems*. Expert Systems with Applications. Vol.34, No.2, February 2008, pp.1445~1455(11).

Lin CY and Hovy E. Identify Topics by Position. 1997. *Proceedings of the 5th Conference on Applied Natural Language Processing*.

Marti A. Hearst and Xerox Palo. 1993. *Automatic Acquisition of Hyponyms from Large Text Corpora*. Proceedings of the Fourteenth International Conference on Computational Linguistics, Nantes France, July 1992M., Debray, S. K., and Peterson, L. L.

Moustafa M. Ghanem, Yike Guo, Huma Lodhi, Yong Zhang. 2002. *Automatic Scientific Text Classification Using Local Patterns*. KDD CUP 2002 (Task 1).

P. Kiran Mayee, Rajeev Sangal, Soma Paul. 2008. *PurposeNet: An Ontological Resource organized around around purpose*. Proceedings of ICON2008: 6th International Conference on Natural Language Processing Macmillan Publishers, India.

Developing a Paradigm Based Morph Analyser for Kashmiri

Aadil Amin Kak   Nazima Mehdi   Aadil Ahmad Lawaye   Feroz Ahmad Lone
Department of Linguistics, University of Kashmir

Abstract
The present paper is based on a paradigmatic approach. The notion of a paradigm being closely related to that of inflection, the paradigm of a word (root) will be the set of all of its forms, organized by their grammatical features. The basic idea behind the paper is that if one paradigm can be identified a common analysis for all the words sharing that paradigm will be evolved till the whole language is covered.

1 Introduction

The role of morphology is very significant in the field of NLP, as seen in applications like MT, IE, IR, Spell Checker, Lexicography, etc. The basic function of a Morphological Analyzer is taking a word as an input and analyzing its grammatical features like its root, stem, affixes and so on.

The word order of Kashmiri is quite different from Indo-Aryan languages. It is a verb second language (i.e. V-2). The word order resembles languages like German, Dutch etc. and is quite different from verb middle languages like English. In a V-2 language, any constituent of a sentence can precede the verb.

Morphologically speaking, words in Kashmiri are either mono or poly-morphic. Grammatical categories can be divided into three main groups on the basis of their morphological characteristics viz:

(Group I).Nouns, Pronouns, Adjectives and Quantifiers fall in one group (Group II). Verbs constitute another group (Group III).Adverbs, Postpositions, Conjunctions, Interjections, and Negative particles constitute another group

The word structure of Kashmiri can be described in terms of the classes of morphemes. A large number of words in Kashmiri are constituted by means of affixes which are bound forms. When these affixes are added to free forms, either verbal or nominal, their senses are realized. Taking these features into consideration an attempt was made to develop a paradigm based morph analyser for Kashmiri.

2 Working

A morphological analyzer is a computational tool, which performs automatic morphological analysis and synthesis of word forms using an electronic dictionary of base forms. A morphological analyzer takes a word as an input and produces the root and its grammatical features as output. Words in the input text are first processed by the morphological analyzer. Its task is to identify the root, lexical category, and other features of the given word. Different Morph analysers are built for different languages like Hindi, Marathi, Sanskrit, Telugu, Kannada and many other languages using different approaches viz; paradigm based and rule based approaches. Keeping these morph analysers into consideration work is going on to develop a Morph analyser for Kashmiri. The morph analyser in question is based on paradigmatic approach. The notion of a paradigm being closely related to that of inflection, the paradigm of a word (root) will be the set of all of its forms, organized by their grammatical features. Many words share paradigms, i.e. more than one word can have same inflectional behavior. In other words we can say if one paradigm can be identified a

common analysis for all the words sharing that paradigm will be evolved till the whole language is covered.

To develop a morph analyser for Kashmiri first Nouns of the language were taken into consideration for which the paradigms were built as follows:

Screenshot 1: Snap shot of Noun Paradigm (Masculine)

In the given snap shot, paradigm for the word *aasmaan* 'sky' is built. The paradigm for the word *aasmaan* is built taking into consideration its all grammatical features like gender, number, and case. The inflectional table of this particular word is given as under.

Cases	Singular	Plural
Nom.	Ø	Ø
Dat.	-as	-an
Erg.	-an	-av
Gen.	-uk	-Ik'
	-Ich	-chi
Abl.	-I	-av

Table 1: Inflectional table of *aasmaan*

Taking these inflections into consideration paradigm for *aasmaan* is built and the words sharing these features will be taken in the same paradigm in the foot note dictionary and will be analysed in the same way as that of *aasmaan*. In this way paradigm for all these words is built.

Screen shot2: Snap shot showing Noun Paradigm (Feminine)

In the given snap shot paradigm for the word '*v :r*' is built. The paradigm for the word v :r is built taking its all grammatical features like gender, number, and case into consideration. The inflectional table of this particular word is given as under.

Cases	Singular	Plural
Nom.	Ø	Ø
Dat.	- i	- 'an
Erg.	- i	- 'av
Gen.	- i-hund	- i-hInd'
	--i-hInz	- i-hInzI
Abl.	-i	- 'av

Table 2: Inflectional table of *v :r*

Taking these inflections into consideration paradigm for *v :r* is built and the words sharing these features will be taken in the same paradigm in the foot note dictionary and will be analyzed in the same way as that of *v :r*. In this way paradigm for all these words is built.

3 Issues and solutions

Like this paradigms for nouns were built and approximately 30 paradigms for nouns were built till now. However there were certain problems faced while using this approach (Paradigmatic approach). The main problem was faced whenever a new word was given to the morph analyser it was not able to analyze that particular word. In order to overcome such an issue a rule based approach was followed where in words are analysed on the basis of the rules of the particular language.

In a rule based approach decision tree is followed to develop the rules and in this way the paradigms are generated by the morph analyser on its own on the basis of rules.

The notion of rule based approach is related to the overall morphological analysis of a particular word. Rules are built on the basis of the grammatical features like Gender, Number, Case and Animacy of a word. Moreover Consonant-Vowel-construction like CVC, CVCVC, CCVC etc is also taken into consideration as Gender, Number and Case markers vary with the Consonant-Vowel construction of a particular word i.e. every word takes different inflections depending upon the consonant vowel construction within the word.

Rules are implemented using Decision tree as shown in the figure given below.
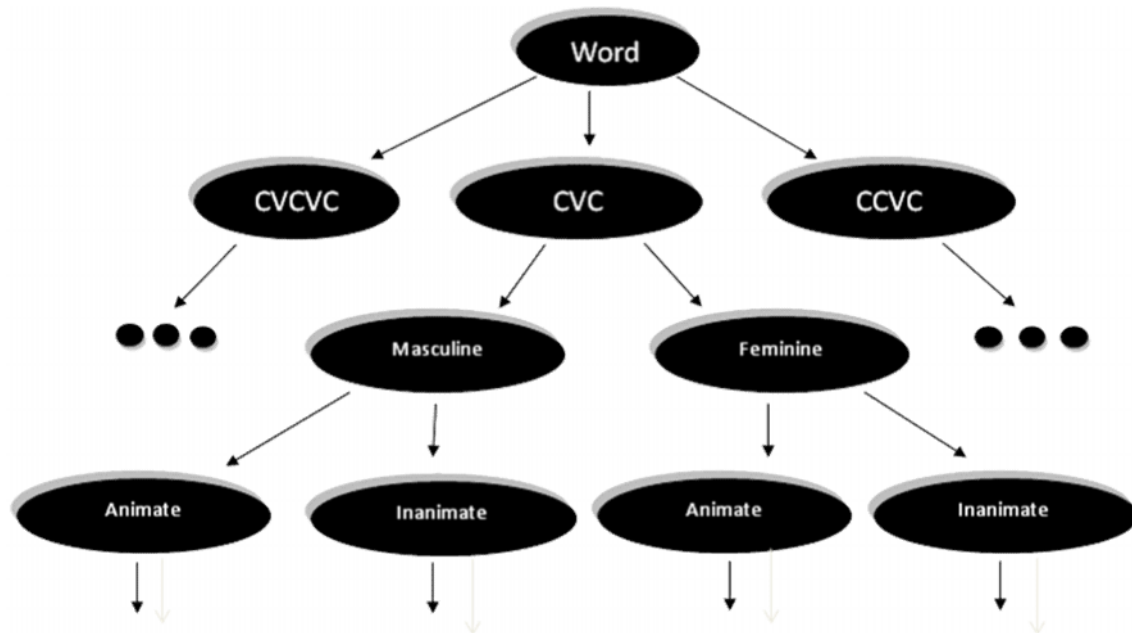
Fig. 1: Decision Tree Implementation of Morph Analyzer.

Taking the morphological richness of the language into consideration, rules are built on the basis of consonant vowel construction. In this scheme, the words sharing the same rule i.e. having the same inflectional behavior will fall under one category. E.g. if we take the example of a CVC construction with medial vowel '**a:**' into consideration, the words having this construction take the following inflections.

Cases	Masculine				Feminine			
	Animate		Inanimate		Animate		Inanimate	
	Singular	Plural	Singular	Plural	Singular	Plural	Singular	Plural
Nom.	Ø	Ø	Ø	Ø	Ø	-I	Ø	-I
Dat.	-as	-an	-as	-an	-i	-an	-i	-an
Erg.	-an	-av	-an	-av	-i	-av	-i	-av
Gen.	I_sund, I_sInz	-an_hund, -an_hInd'	-uk, -Ich	-Ik', -chi	-i_hund, -i_hInd'	-an_hund, -an_hInd'	-i_hund, -i_hInd'	-an_hund, -an_hInd'
	-I_sInd', -I_sInzI	-an_hInz, -an_hInzI			-i_hInz, i_hInzI	-an_hInz, -an_hInzI	-i_hInz, i_hInzI	-an_hInz, -an_hInzI
Abl.	-I	-av	-I	-av	-i	-av	-i	-av

Table 3: Inflectional behavior of CVC (medial vowel 'a:') Construction.

On the basis of the given inflections one set of rules will be formulated and the words having the same construction i.e. CVC (with the same vowel) will fall under the same rule. For exceptions, if any, separate rules will be formulated on the basis of ultimate and penultimate consonants of the last syllable of the word.

In this way the approach used in the present paper is a hybrid approach where the core form is based on paradigms and supported by rules when and if needed. Till now rules for Nouns have been completed and the work is going on other parts of speech as well.

4 Conclusion

The paper in question was based on paradigmatic approach in which paradigms (30 paradigms) for nouns were built and work is also going on other part of speech as well. In the present paper hybrid approach is used to develop a morph analyser where rules were built to generate paradigms and also work is going on to develop a sizeable lexicon which will be given to the morph analyser

to authenticate the rules. Till now rules for nouns are almost complete with some exceptions of course and work is also going on the analysis part of the morph analyser with the help of which the analyser will be able to analyse a word if given with its inflections

Reference

Bharati, A., V. Chatanya, and R. Sangal. 1995.*Natural Language Processing: A Paninian Perspective*. New Delhi: Prentice Hall,.

Bhat, R. K. 1987 *A Descriptive Study of Kashmiri*. New Delhi: Amar Prakashan.

Jurafsky, D., and H. J. Martin. 2000.*Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. London: Prentice Hall.

Kachru, B. B. 1969.*A Reference Grammar of Kashmiri*. Urbana-Champagne: U of Illinois Press.

Wali, K., and O. N. Koul.1997 *Kashmiri: A Cognitive-Descriptive Grammar*. London: Routledge.

Role of working memory in typically developing children's complex sentence comprehension

Ms. Trupthi T.   Ms. Deepthi M.   Ms. Shwetha M.P   Ms. Deepa M .S
rao.trupthi@gmail.com   deepthi_snh@yahoo.co.in   shwth.jain@gmail.com   deepams12@gmail.com
Dept. of Speech-Language Pathology, J.S.S. Institute of Speech and Hearing, Ooty Road,Mysore
.

ABSTRACT

Working memory refers to the ability to store information while at the same time engage in some kind of cognitively demanding activity. These include a phonological short-term memory (PSTM) storage buffer and a visuo-spatial short-term memory buffer, an attentional resource control function, and processing speed. The ability to repeat nonwords is considered a sensitive index of PSTM capacity. We investigated the influence of phonological short-term memory, attentional resource control/allocation on children's complex (and simple) sentence comprehension. Forty children divided into group I (7-9yrs) and group II (10-12 yrs) who completed a nonword repetition task and concurrent verbal processing-storage task. Statistical analysis was done on data obtained and results revealed that (1) children were significantly more accurate in repeating short nonwords than longer nonwords reflecting the capacity-limited nature of the PSTM buffer (2) The memory variables didn't correlate with simple sentence comprehension, (3) resource control/allocation correlated significantly with complex sentence comprehension, even after co-varying for age. In concurrent processing-storage task children comprehended significantly more simple sentences than complex sentences along with production of nonwords. In sentence comprehension task group II performed better than the group I for complex sentence comprehension. It appears that controlled/flexible use of attention plays a major role in complex sentence comprehension.

1 Introduction

Working memory refers to the ability to store information while at the same time engage in some kind of cognitively demanding activity such as verbal reasoning or comprehension (e.g., Baddeley 1999, 2000).

Baddeley and Hitch (1974) introduced the multi-component model of working memory. This theory proposes that two "slave systems" are responsible for short-term maintenance of information, and a "central executive" is responsible for the supervision of information integration and for coordinating the slave systems.

One slave system, the phonological loop, stores phonological information (i.e., the sound of language) and prevents its decay by continuously articulating its contents, thereby refreshing the information in a rehearsal loop. The other slave system, the visuo-spatial sketch pad, stores visual and spatial information. It can be used, for example, for constructing and manipulating visual images, and for the representation of mental maps. The central executive is, among other things, responsible for directing attention to relevant information, suppressing irrelevant information and inappropriate actions, and for coordinating cognitive processes when more than one task must be done at the same time.

Many developmental memory researchers like Bayliss et al. 2005, Conlin et al. 2005, Gathercole 1999, Gathercole et al. 2004, Towse et al. 1998, 2002 conceptualize working memory as a

multidimensional system comprising several interactive, inter-related mechanisms. These include a phonological short-term memory (PSTM) storage buffer and a visuo-spatial short-term memory buffer, an attentional resource control function, and processing speed. A growing body of evidence supports the existence of these mechanisms.

Phonological short term memory and attentional resource control/allocation mechanism and their potential association with complex sentence comprehension in typically developing (TD) children between the ages of 7 and 12 years were the focus of the present study.

1.1 Phonological Short-Term Memory

The ability to repeat nonwords is considered a sensitive index of PSTM capacity because successful repetition requires children to invoke a variety of phonological and memory-related processes (perception, encoding, storage, retrieval, production) independent of lexical knowledge (Gathercole and Baddeley 1990, Henry and Millar 1991). This is especially the case when the stimuli do not closely resemble real words in their canonical stress pattern or phonotactic probability of occurrence (Dollaghan et al. 1993, Gathercole et al. 1999, Hulme et al. 1997. Non-word repetition has shown itself to be a robust index of children's PSTM capacity (Gathercole 1999).

1.2 Attentional Resource Control and Allocation

The working memory system is considered to have limited attentional capacity that can be allocated to support information processing or storage or both (e.g., Daneman and Carpenter 1980, Just and Carpenter 1992). Resource capacity refers to the amount of cognitive capacity (i.e., mental energy) available to an individual to perform a mental task. Resource control (e.g., one function of the central executive according to Baddeley 1999) refers to a person's ability to flexibly control and divide his/her attentional resources between two or more different concurrent mental activities. Children's ability to engage in simultaneous information processing and storage improves with age, particularly during the school age and early adolescent years (Barrouillet and Camos 2001, Gathercole 1999, Gaulin and Campbell 1994, Gavens and Barrouillet 2004, Johnson 1997).

Using similar tasks other researchers have shown that young school age children have the ability to allocate their attentional resources simultaneously to verbal processing and storage (Marton and Schwartz 2003, Montgomery 2000a,b). Marton and Schwartz, asked children to listen to sentences that contained a nonword (2-, 3- or 4-syllable nonword) at the end of short-simple, short-complex, and long-complex sentences. After each sentence, children were asked to repeat the nonword (indexing PSTM capacity) and answer a Y/N question (indexing concurrent language processing). These investigators reported that children showed a strong tendency to mispronounce the non-words (regardless of syllable length) and miscomprehend sentences as the sentences became more complex. For the simple sentences, however, the children showed considerably better non-word production and comprehension. Such findings indicate that children clearly have the ability to allocate and coordinate their attentional resources to both verbal processing and storage but show a performance (i.e., verbal storage) decrement as the language processing demands reach or exceed the limits of their attentional capacity.

Majority of childhood memory-language research has focused on the association between working memory, especially the PSTM buffer, and lexical learning (e.g., Gathercole and Baddeley 1990, Gathercole et al. 1997, 1999). By contrast, relatively few studies have investigated the relation between working memory and sentence comprehension, with lesser number of studies designed specifically to examine its influence on complex sentence comprehension

There exits some research examining the relation between working memory and children's general sentence comprehension (e.g., Adams et al. 1999, Ellis Weismer and Thordardottir 2002,

Montgomery 1995b, 2000a,b). Montgomery (1995b), explored this interaction of working memory and general sentence comprehension, the school age typically developing children (and children with language learning impairment) to complete a nonword repetition task and a sentence comprehension task. Results of a correlation analysis revealed a strong positive association between nonword repetition and comprehension, suggesting that PSTM capacity is important to children's sentence comprehension. But were not able to hypothesize what is the association is between PSTM capacity and complex sentence comprehension.

Ellis Weismer and Thordardottir (2002) studied the association of both PSTM and resource control/allocation on the general sentence comprehension of a group of five to nine year old typically developing children and children with SLI. Children completed a nonword repetition task (index of PSTM capacity), the CLPT (index of resource control/allocation), and a standardized sentence comprehension test that included a wide range of simple and complex constructions. Between the memory variables, the most unique variance in children's sentence comprehension was accounted for by resource control/allocation, followed by PSTM.

2 Aim of the Study

To our knowledge, there are few researches examining the influence of PSTM, attentional resource control/allocation on children's complex sentence comprehension. The present study therefore was designed examine the influence of working memory on complex structures that are well within children's linguistic competence and experience.

3 Method

3.1 Participants

Forty typically developing children between 7 and 12 years of age participated in the study. They were divided into two groups, group I (7-9 years) and group II(10-12 years) with 20 children in each age group, with equal number of males and females (10 males and 10 females).Children in this age range were studied because this is the age range in which reliable correlations have been reported between one or more of these memory variables and lexical learning complex memory performance and short-term memory span, and academic skill development

3.2 Selection criteria

All the children demonstrated
1. Normal-range hearing sensitivity as determined by audiometric pure tone screening at 25 dB HL (ANSI, 1990) for 500 Hz, 1 kHz,2 kHz, and 4 kHz.
2. Average or above average academic performance [from teacher report].
3. Kannada as their primary spoken language.
4. No oral structural/motor impairments affecting speech or non-speech movements of the articulators
5. Children had no history of neurological impairment or psychological/emotional disturbance or attention deficit disorder (from parent report).

3.3 Stimulus Generation Procedures for the Verbal Memory and Comprehension Tasks

Procedures for recording, generating, and editing the verbal stimuli for the memory and sentence comprehension tasks were identical. A Female speaker with Kannada dialect sat in an isolated acoustic booth and read the stimuli using a high quality microphone connected to a Dell PC. The nonwords were read in list fashion. The sentences for the resource allocation task and sentence

comprehension task were read at a normal rate (~4.4 syllables/sec) and with normal prosodic variation. All stimuli were digitized at the sampling rate 16 kHz.

3.4 Phonological Short-Term Memory (PSTM): Non-word Repetition Task

Task Design and Stimuli

Each child completed a conventional nonword repetition (NWR) task as the index of PSTM capacity. The task, which was created for present study, included 20 nonwords varying in length from 2 syllable to 5 syllables. Stimulus words were of CVC, CVCVC-type structure and followed kannada phonotactic rules. Stimulus words did not contain any word-initial consonant clusters,which were designed to minimize complex articulatory/output demands

Procedure

Children were instructed to listen to some "funny"/"pretend" meaningless words and to repeat each one immediately after hearing it. In keeping with conventional administration practices children received one presentation of each item. Nonwords were output from the computer (16 kHz). Each nonword was scored live by the examiner as either + or −. A speech error was defined as any phoneme deletion, substitution, or addition. Consistent phoneme distortions were not counted wrong. The dependent variable was percent nonwords produced correctly at each syllable length. Example stimuli appear in Appendix A.

3.5 Attentional Resource Control/Allocation: Concurrent Processing-Storage Task

A concurrent verbal processing-storage (CPS) task was used as the index of attentional resource control/allocation .we regard this task to reflect children's general use of controlled and flexible attentional abilities, i.e., allocation of attention to the language processing system and PSTM buffer. The task includes sentences that will be well within the linguistic competence of all the children any difference between the younger and older children can be interpreted to be attributable to age-related differences in resource allocation, not linguistic knowledge.

Task Design and Stimuli

The task included 10 simple sentences (e.g |hakkiju| |mele| |harutha| *|ma:tsire|*) and 10 sentences with a center embedded NP-modifier clause (e.g., |bharavaseja| |a:taga:rana:gi| |belejuthiruva| |bada| |huduganu| *|sovalige|*). Simple and longer/complex sentences were used to vary the linguistic processing demands of the task. Even though the children will have considerable experience with each of these sentence types, the longer/complex sentences require greater attentional resources than the simple sentences (Marton and Schwartz 2003). All of the sentences contained highly familiar nouns and verbs to 7 year old children.

Simple sentences contained four words and the longer/complex sentences contained six to seven words. The last word in each sentence contained either a 3- or 4-syllable nonword. For each sentence type, an equal number of 3-syllable or 4-syllable nonwords appeared at the end of the sentence. Two-syllable nonwords were not included because extensive piloting showed a ceiling effect, with children performing exceptionally well. Five syllable nonwords were also not included because the same pilot testing showed a floor effect.

Following the same constraints as in the nonword repetition task, the nonwords in this task were also carefully controlled for their phonotactic and prosodic features. For each stimulus sentence a yes/no question was created. Across simple and complex sentences approximately equal number of questions requiring a "yes"/"no" response was created (e.g. |huduganu| |Shrimantane|?).

Procedure

Children were told they would hear some sentences and that the last word of each sentence was a funny/made-up meaningless word. Immediately following the end of the sentence children repeated the nonword and then answered "yes" or "no" to the comprehension question, which appeared 3 seconds after the nonword. Pilot testing showed that this interval was sufficient for children to repeat the nonword before hearing the question. Each stimulus sentence was presented just once.Sentences were output from the computer (16 kHz) and Children heard two demonstrations. For each experimental trial the examiner transcribed live the child's nonword production and recorded his/her answer to the question on his/her individual score sheet. The task also was recorded on the digital recorder for later broad phonetic transcription and reliability checking. A speech error was defined as any phoneme deletion, substitution, or addition. Consistent phoneme distortions were not counted wrong. The primary dependent variable was the percent of sentences for which both the nonword was produced correctly and the comprehension question answered correctly. Example items appear in the Appendix.

4 Results and Discussion

Children's Performance on the Individual Experimental Tasks

4.1 Phonological Short-Term Memory: Nonword Repetition Task
On the nonword repetition (NWR) task, statistical analysis revealed a significant syllable length effect. As seen in Graph 1, results of means of samples showed that the children correctly repeated comparable numbers of 2-syllable items than 3-syllable items. They did repeat fewer numbers of 4-syllable items correctly relative to 3-syllable items and significantly fewer 5-syllable items compared with 4-syllable items. There was significant poorer performance seen in the repetition of 5-syllable word in both the groups. There was drastic decrement in the performance for 5-syllable repetition task in group I compared to group II as seen in graph I. This can be attributed to the poorer performance by the younger children than the older ones when the phonological short term memory demand increased.
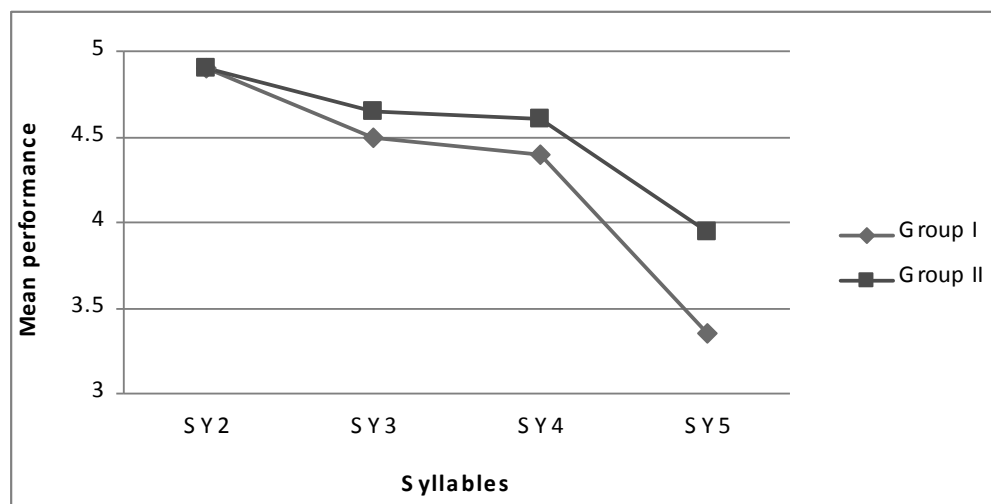
(sy2 = 2syllable words, sy3= 3 syllable words, sy4= 4 syllable words, sy5= 5 syllable words)
Graph1: Performance of group I and group II in nonword repetition task with increasing syllable length

118

This is in support of studies done by Dollaghan and Campbell 1998, Dollaghan et al. 1993, Ellis Weismer et al. 2000, Gathercole 1999, Gathercole and Baddeley 1990, Montgomery 2004. The children were significantly more accurate repeating short nonwords than longer nonwords. Such a pattern has been interpreted by many investigators to reflect the capacity-limited nature of the PSTM buffer (Baddeley 2003, Gathercole and Baddeley 1990, Montgomery 2004).When performance was compared between groups, group II performed better than group I, though it was not statistically significant. Comparision between the genders within the groups, females in group II performed significantly better than males.

Attentional Resource Control/Allocation: Concurrent Processing-Storage Task
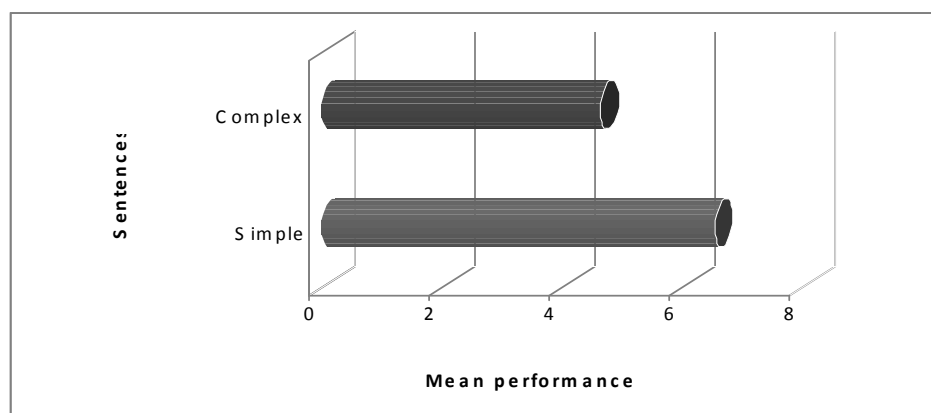
Graph2. P*erformance of children in Concurrent Processing-Storage Task with complex and simple sentences*

Performance on the concurrent processing-storage (CPS) task was evaluated using a sentence type (simple, complex) with non-word. The primary dependent variable was the percent of sentences for which both the nonword was accurately repeated and the sentence correctly comprehended. Results revealed a significant sentence type effect with non word repetition in both the groups.

In Graph2, it can be seen that children comprehended significantly more simple sentences than complex sentences along with production of nonwords. In sentence comprehension task group II performed better than the group I for complex sentence comprehension and within the groups there was gender difference. In Group II the performance of males was better in complex sentence comprehension. Among both the groups and gender the performance of males in group II was best when compared to females of same group and both the gender of other group. The children's performance on the Concurrent processing-storage task was in good agreement with previously published reports showing that school age children have the ability to allocate their attentional resources simultaneously to verbal processing and verbal storage (Ellis Weismer et al. 1999, Gavens and Barrouillet 2004, Marton and Schwartz 2003, Montgomery 2000a,b).Children achieved a mean score of 65% correct, reflecting both accurate nonword production and correct sentence comprehension of simple sentences. While children showed good ability to divide their resources concurrently to both the PSTM buffer and to the language processing system they also showed a processing-storage balance (e.g., Just and Carpenter 1992, Ellis Weismer et al. 1999, Gaulin and Campbell 1994, Marton and Schwartz 2003), as evidenced by the significant sentence type with nonword. The interaction was reflected in the fact that the children yielded significantly poorer nonword repetition when processing longer/complex sentences relative to simple sentences. The

memory measures were found to correlate with age, findings that agree well with the developmental memory literature.Memory variable did not correlate with children's simple sentence comprehension. For simple and short sentences children's working memory system involvement is not expected. Comprehension accuracy for complex sentences was significantly associated with attentional resource control/allocation .These associations remained significant even after ruling out the effects of age. The Concurrent-processing storage task score that is non word repetition and comprehension, with the fact that 25% of the variance in comprehension between complex and simple sentence was accounted for by the Concurrent-processing storage score, indicates that children's comprehension of highly familiar complex sentence structures still requires considerable controlled attentional functioning.
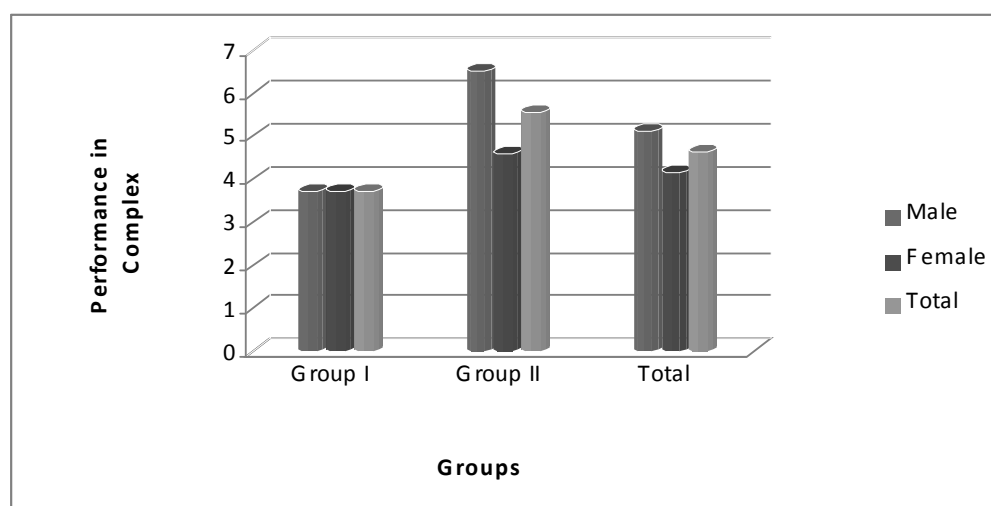
Graph3.Performance of children in Concurrent Processing-Storage Task across group and gender in complex sentences.

The findings demonstrate a clear association between working memory and complex sentence comprehension in children of this age range. Our findings are also consistent with those reported in the adult language processing literature in that relative to simple sentences complex structures give rise to greater working memory involvement. The adult literature has shown significant correlations between performance on various concurrent processing-storage tasks and complex sentence comprehension measures (King and Just 1991).

5 Conclusion

The present study was designed to examine the influence of working memory on complex structures. We considered 40 children in the age range 7-12 years and were divided into group I(7-9yrs) and group II(10-12 yrs) . The phonological short term memory was assessed with the task of non-word repetition task and attentional resource control and allocation was assessed using concurrent processing storage task of simultaneous non word repetition and comprehension for simple sentence and complex sentence. The children were significantly more accurate repeating short nonwords (e.g., 2syllable items) than longer nonwords (e.g., 3-, 4- and 5-syllable items). Such a pattern reflects the capacity-limited nature of the PSTM buffer. Group II performed better indicating the reduced access to the working memory in group I.

In concurrent processing-storage task children comprehended significantly more simple sentences than complex sentences along with production of nonwords. In sentence comprehension task group

II performed better than the group I for complex sentence comprehension. In phonological short term memory the females in group II performed better whereas in attentional resource control and allocation males of group II performed better. It appears that controlled/flexible use of attention plays a major role in complex sentence comprehension.

Because this study was the first of its kind, the present findings would be supported if they could be replicated. Future studies might also investigate the differential role of the different working memory mechanisms on a greater range of complex (and longer) sentence structures (e.g., various relative clause structures). Future investigations may also wish to take a developmental approach by examining whether different associations might hold between the various working memory mechanisms and complex sentence comprehension at different ages.

Acknowledgment

REFERENCES

Adams, A., Bourke, L., &Willis, C. 1999.*Working memory and spoken language comprehension in youngchildren. I*nternational Journal of Psychology, 34, 364–373.

Baddeley, A. 2003.*Working memory and language: An overview.* Journal of Communication Disorders, 36,189–208.

Barrouillet, P., & Camos, V. 2001. *Developmental increase in working memory span: Resource sharing or temporal decay?* Journal of Memory and Language, 45, 1–20.

Daneman, M., & Carpenter, P. 1983. *Individual differences in integrating information between and withinsentences.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 561–584.

DeVilliers, J., & DeVilliers, P. 1973. *Development of the use of word order in comprehension.* Journal of Psycholinguistic Research, 2, 331–342.

Dollaghan, C., & Campbell, T. 1998. *Nonword repetition and child language impairment*. Journal of Speech, Language, and Hearing Research, 41, 1136–1146.

EllisWeismer, S., & Thordardottir, E. 2002. *Cognition and language. In P. Accardo, B. Rogers,&* A. Capute (Eds.), Disorders of language development. (pp. 21–37). Timonium, MD: York Press, Inc.

Gathercole, S., Pickering, S., Ambridge, B., & Wearing, H. 2004. *The structure of working*

memory from 4 to 15 years of age. Developmental Psychology, 40, 177–190.

Just, M., & Carpenter, P. 1992. *A capacity theory of comprehension: Individual differences in working memory.* Psychological Review, 99, 122–149.

Montgomery, J. 2000a. *Relation of working memory to off-line and real-time sentence processing in children with specific language impairment*. Applied Psycholinguistics, 21, 117–148.

Montgomery J, Magimairaj BM,O'MalleyMH.2008. *Role of Working Memory in Typically Developing Children's Complex Sentence Comprehension*. Journal of psycholinguist research.

Towse, J., Hitch, G., & Hutton, U. 1998. *A reevaluation of working memory capacity in*

children. Journal of Memory and Language, 39, 195–217.

Vallar, G., & Baddeley, A. 1984. *Phonological short-term store, phonological processing and sentence comprehension:A neuropsychological case study.* Cognitive Neuropsychology, 1, 121– 141.

APPENDIX A: Example stimuli for memory Tasks

Phonological short-term memory (Nonword Repetition task) stimuli

1. |paba|	3.|lafi:na|	5.|savalina|	7.|navasakalja|
2.	duna|	4.|vumaji|	6.|dzhajarika|	8.|rapran.ikali|

Attentional resource capacity/allocation (concurrent processing-storage task) stimuli

Simple sentences
1.|hakkiju| |mele| |harutha| |**ma:tsire**| question: |hak'iju| |ha:rutitha| ?

2.|me:nu| |e:dzutha| |pakakke| |**kulaidine**| question: |me:nu| |ha:rutithe| ?

3.|ha:vu| |tevalutha| |munde| |**nikrushe|**   question: |ha:vu| |malagithe|?

4.|nariju| |negejutha| |pakakke| |**latnugere|**   question: |nariju| |negejutithe|?

5.|mi:nu| |i:dzuta| |me:lake| |**kulaidine**| question: |mi:nu| |harutithe|?

1.1.2   *complex sentences*
1.|bh'aravaseja| |a:taga:rana:gi| |belejuthiruva| question: |hud.uganu| |Shrimanthane|?
  |bada| |huduganu| |**sovalige|.**

2.|ha:ruthidha| |bannada| |ga:lipata| question: |ga:lipat.a| |bannabannadagite|?
  |marakke| |sikkihakikondu| |**orjakodze|**

3.|kottigejalidda| |bili| |hasuvu| |hakida| question: |hasuvu| |bajalinalithe|?
  |mevannu| |tinutta| **|ma:chidakshe|**

4. |hasididda| |balakanu| |maradinda| question: |ba:lakanu| |hasidid'ane|?
  |dzora:gi| |kelage| |dzigidu| **|thukelagna|**

5. |te:luthiruva| |kappu| |mo:dagalu| question: |mo:da| |bilijada:githe|?
  |male| |surisutha| **|lidzigrave|**

Pronominal Binding in Hindi-Urdu vis-à-vis Bangla

Aparna Mukherjee
Centre for Linguistics
Jawaharlal Nehru University
New Delhi, India
aparna.jnu27@gmail.com

Abstract

The paper aims to study the syntactic properties of possessive pronominals in a language like Hindi-Urdu, in contrast to Bangla. Hindi-Urdu unlike Bangla exhibits a kind of pronominal binding in which the possessive pronominal must obligatorily obviate from the closest c-commanding subject antecedent. This behaviour[1] is not captured by the Principle B of the standard binding theory.

1. Introduction

Hindi-Urdu unlike Bangla exhibits a kind of pronominal binding in which the possessive pronominal must obligatorily obviate from the closest c-commanding subject antecedent. Hindi-Urdu has reflexive possessive that is distinct from the pronominal possessives. The possessive pronominals are anti-subject oriented while the possessive reflexives are bound by the subject. Hindi-Urdu possessive pronominals behave similar to those in languages like Danish (Vikner, 1985), Norwegian (Hestvik, 1990, 1992), Russian (Avrutin, 1994). In Bangla the situation is reminiscent of that in German, Dutch and English. It has no specialised reflexive and the possessive pronominals are bound by the subject.

The binding theory designed to show the facts for English and Bangla type languages alone cannot explain why the Hindi-Urdu type languages show a distinct pattern. In order to learn the nature of this contrast in languages like Hindi and Bangla, we have to be clear in theoretical perspectives. The Canonical Binding Theory (Chomsky 1981) treated pronominals and anaphors as the properties of primitives of the system and the notion of co-indexing to represent the anaphoric dependencies, to a large extent.

As discussed in Chomsky (1995), while explaining the foundations of the properties of the Minimalist Programme, the role of indices in the CBT is incompatible with the inclusiveness condition that limits the syntactic computations to operations on the basis of a purely morpho-syntactic features of the lexical item and prevents the introduction of the elements in the course of their derivation. Syntactic operations such as Agree and semantic component of binding have to develop syntactic dependencies. On the basis of this the un-interpretable features get deleted but must be visible in the morphology.

2. Hindi-Urdu

Possessive pronominals in Hindi-Urdu as in sentences 1, 2 and 3 must obligatorily obviate from the closest c-commanding subject antecedent. The embedded possessive pronominal may be bound by a higher c-commanding object or by a subject outside its containing clause, as in sentence 2. However, non-embedded pronominals, as in sentence 13, respect the Principle B of the standard

---

[1]   The PP embedded pronominal behaves very similar to the possessive pronominals.

binding theory that imposes a condition of referential independence on pronominals in their binding domain, the CFC. Outside this domain they may freely be used co-referentially.

Sentences with Possessive pronominals:

1 ra:m$_i$ us-ke$_{*i/j}$ k mr-e   me˜   g -y-a:
  ram   3-Gen   room-Obl   in   go-Past-Mas
  'Ram$_i$ went into his$_{*i/j}$ room'

2 ra:m$_i$ ne   moh n$_j$ ko   us-ki$_{*i/j}$ kita:b di:
  ram Erg mohan Acc   3-Gen-Fem   book   give-Past-Fem
  'Ram$_i$ gave his$_{*i/j}$ book to Mohan$_j$'

3 h r   ma˜:$_i$ us-ke$_{*i/j}$ b cce se   pya:r   k r-t-i
  every   mother 3-Gen-Mas child with love   do-Hab-Fem

  h
  be-Pres
  'Every mother$_i$ loves her$_{*i/j}$ child'

Sentences with Possessive Reflexives/ Anaphors:

4 ra:m$_i$ pne$_i$ k mr-e   me˜   g -y-a:
  ram   self-Gen   room-Obl   in   go-Past-Mas
  'Ram$_i$ went into his$_i$ room'

5 ra:m$_i$ ne   moh n$_j$ ko   pn-i$_{i/*j}$ kita:b di:
  ram Erg   mohan Acc   self-Gen-Fem book   give-Past-Fem
  'Ram$_i$ gave his$_{i/*j}$ book to Mohan$_j$'

6 h r   ma˜:$_i$ pne$_i$ b cce se   pya:r   k r-t-i
  every   mother 3-Gen-Mas   child with love   do-Hab-Fem

  h
  be-Pres
  'Every mother$_i$ loves her$_{i/*j}$ child'

Kidwai (2000) has explained this anti-subject orientation of the possessive and PP pronominals by an extension of Avrutin's (1994) LF raising approach to reflexive interpretation to pronominals. Under her proposal, in Hindi-Urdu the pronominals need to be raised at LF to a functional projection of D^0 unlike those in English type languages where the structural requirement is satisfied at the Spellout.

However in the current assumptions this phenomenon needs further explanation. In this paper the anti-subject orientation of the possessive pronominals in Hindi-Urdu has been looked at by making a comparison between the syntactic dependency of the pronominal and the reflexive on the [Spec, TP], the landing site of the Subject. This means that such expressions are interpreted in terms of some other expressions instead of assigning value to them. This dependency can be explained on the basis of the feature determinacy thesis. (Reuland, "Anaphora in Language Design").

Feature Determinacy Thesis

Syntactic binding of pronominal elements (including 'anaphors') in a particular environment is determined by their morpho-syntactic features and the way these enter into the syntactic operations available in that environment.

(Reuland, "Anaphora in Language Design")

This thesis shifts the focus in the investigation of binding from macro principles such as the canonical binding condition, to the question of what types of feature clusters allow or enforce the formation of syntactic dependency under different conditions. This syntactic dependency, is established by Move/Attract and Checking, and it is forced by checking grammatical features in a checking configuration. If the features are in a checking configuration, it entails that the features are subject to deletion or erasure that can be recovered. This is the Principle of Recoverability of Deletion (PRD).

The morpho-syntactic features of the pronominals are [+/-Gender], [+Number], [+Person] while that of the reflexives and other anaphors are [-Gender], [+/-Number], [-Person]. The anaphors except the pronominals are syntactically dependent on the [Spec, TP] to fulfil their feature value. The anaphors are locally bound by the subject in the local domain where the two positions are transparent and hence visible to each other while the possessive pronominals need not be bound though they may occur in the c-commanding domain of the subject.

3. Bangla

Bangla pronominals behave like English pronominals. As can be seen in the sentences 7, 8 and 9, the pronominals embedded in DP and PP constituents can be co-indexed with the closest subject or with a higher c-commanding object or with a subject outside its containing clause. But the non-embedded ones do not allow such a co-referential use with other co-referential expressions within a sentence. The Bangla DP structure as proposed by Bhattacharya (1998) has functional projections for Dem. According to this proposal the Poss is generated at a much lower shell and moved to a Case position, that is the Spec DP.

The pronominal is in a position where the syntactic dependency on the subject cannot be licensed. The pronominal binding is taken care of by the semantics, where it is co-valued by the subject, either by co-reference or by variable binding.

The definite marker in Bangla, that is the Classifier, is potential enough to cause opacity to allow chain formation. Syntax cannot see through the system hence semantics decides the binding of the possessive pronominal in Bangla.

Sentences with Possessive pronominals:

7 ram$_i$ o-r/ta-r$_{i/j}$ ba i   gæ-l-o
  ram   3N-Gen   room-Loc   go-Past-3N
  'Ram$_i$ went into his$_{i/j}$ room'

8 ram$_i$ mohon$_j$ ke   o-r/ta:-r$_{i/j}$ boi   di-l-o
  ram   mohan Acc   3N-Gen   book give-Pst-3N
  'Ram$_i$ gave his$_{i/j}$ book to Mohan$_j$'

9 prottek- i   ma˜$_i$ o-r/ta-r$_{i/j}$ ba a ke   b^{h}alobaš-e
  every-Clsfr mother 3N-Gen child Acc   love-Hab-3N
  'Every mother$_i$ loves her$_{*i/j}$ child'

Sentences with Possessive Reflexives/ Anaphors:

10 ram$_i$ nij-er$_i$ ba i   gæ-l-o
  ram   own-Gen   room-Loc   go-Past-3N

'Ram$_i$ went into his$_i$ room'

11 ram$_i$ mohon$_j$ ke   nij-er$_{i/*j}$ boi   di-l-o
   ram   mohan Acc   own-Gen   book   give-Pst-3N
   'Ram$_i$ gave his$_{i/*j}$ book to Mohan$_j$'
12 prottek- i   ma˜$_i$ nij-er$_i$ ba a ke   b^{h}alobaš-e
    every-Clsfr mother own-Gen   child Acc   love-Hab-3N
    'Every mother$_i$ loves her$_i$ child'

4. Apparent problem

The Bangla 'nij', which is a prima facie case of anaphor as in Hindi-Urdu, behaves more like the English 'own' that modifies the possessive pronominal.

13. ram$_i$ nij-er$_i$ ba i   gæ -l-o
    ram   own-Gen   home   go-Pst-3.N

14. ram$_i$ ta-r$_i$ nij-er$_i$ ba i   gæ -l-o
    ram   3.SEQ-Gen   own-Gen   home   go-Pst-3.N
    Ram$_i$ went to his$_i$ (own) home.

The pronominal can be dropped as in sentence 15 or it can optionally occur as in sentence 16. Though the latter is less acceptable than the former, when the embedded DP is topicalized, as shown in sentences 17 and 18, the sentence with the POSS pronominal becomes more acceptable.

15. nij-er$_i$ ba i   ram$_i$ gæ -l-o
    own-Gen   home   ram   go-Pst-3.N

16. ta-r$_i$ nij-er$_i$ ba i   ram$_i$ gæ -l-o
    3.SEN-Gen   own-Gen   home   ram   go-Pst-3.N
    It is his$_i$ own house that Ram$_i$ went to.

Co-occurrence of two anaphors, the reflexive and pronominal is not expected as shown by Bangla 'nij'.
Another evidence to suffice the anomalous behavior of 'nij' is seen in the embedded contexts, where Bangla 'nij' is not locally bound. This can be seen in sentence 19.

17. ram$_i$ mohon$_j$ ke   bol-l-o   je   ami$_k$ ta-r$_{i/j/*k}$ nij-er$_{i/j/*k}$ ba i   r
    ram mohan-ACC tell-Pst-3.N that 1.Sg   3.SEQ-Gen own-Gen house colour

    kor-ec^{h}il-am
    do-Pst.Perf-1
    Ram$_i$ told Mohan$_j$ that I had painted his$_{i/j}$ own$_{i/j}$ house.

5. Conclusion

The behaviour of the possessive pronominals in Hindi-Urdu contrasts to that in Bangla, in that, in the former, the possessive pronominals are anti-subject oriented, but this is not the case in Bangla. The discussion so far shows that the pronominal binding in Hindi-Urdu type languages is syntactically motivated while in Bangla it is at the discourse level. The definiteness marker in Bangla blocks the chain formation of the POSS pronominal with the subject, hence the syntactic dependency as well. The environment of the occurrence of the POSS pronominals in a language decides chain formation which eventually leads to the binding of the entities.

References

Bhattacharya Tanmoy, 1998. *DP-Internal NP Movement.* UCL Working papers in Linguistics 10.

Kidwai Ayesha, 2000. XP-Adjunction in Universal Grammar Scrambling and Binding in Hindi-Urdu, Oxford University Press, 90-96.

Reuland, Eric. 2001. Primitives of Binding. *Linguistic Inquiry* Vol. 32. Number 3. 439-92.

Reuland, Eric. 2008. Anaphoric dependencies: How are they encoded? Towards a derivation-based typology. In *Reciprocals and Reflexives. Theoretical and Typological Explorations*, eds., König, Ekkehard, and Gast, Volker. Berlin: Mouton de Gruyter.

Reuland Eric, forthcoming. Anaphora and Language Design. MIT Press.

Reuland Eric, in press. '*Syntax and Interpretation Systems: How is their labour divided*' in Handbook of Minimalism. Ed. Cedric Boeckx, Cambridge University Press .

Corpus based study of Relative clauses in Hindi and Telugu; Transfer Grammar Rules for Relative clauses

Y Viswanatha Naidu
Language Technologies Research Center
International Institute of Information Technology
Hyderabad, India
vnaidu@research.iiit.ac.in

Abstract

Machine Translation (MT) has gained increasingly more interest since it was developed, but yet it is not reaching the users expectations. It would be improved, using lexical resources like tree-bank, verb-net, word-net, tools like tokenizer, morph-analyzer, generator, parser and also an amount transfer grammar. These resources, tools and grammars play a major role in MT. Among them transfer grammar is an approach to translate the structures of source language (SL) to target language (TL). It is one of the structural grammars which are a central design in MT. The aim of this paper is to study the nature of relative clauses in Hindi and Telugu and to develop the transfer grammar for the relative clauses.

1 Background

Non-corpus based research had already been done on Telugu to see which NP can be relativized in Telugu in terms of the noun phrase accessibility proposed by Keenan and Comrie (1977) (Usha Rani. A, 1990). But the present study focuses on the analysis of relative clauses based on corpus in both Hindi and Telugu languages.[1] Benefit of the corpus-based research is that it can have much broader samples of typical uses of grammar structures. This paper aims to investigate the nature of the relative clauses by extracting the data from POS-tagged and chunked corpus. Then, based on Keenan and Comrie's Accessibility hierarchy[2] (AH), relative clauses of both Hindi and Telugu are studied. Subsequently, efforts are also made to make a set of Transfer Grammar rules, that are required for mapping the syntactic representations of a SL (Hindi relative clauses) to TL (Telugu relative clauses). We know that languages have their own mechanisms to encode the information. Some languages encode it explicitly through linguistic elements which other languages may not express explicitly, in which case they will have their own mechanism to encode the information. For instance Hindi cannot relativize the instruments in sentences like *jis cammac se maine khAyA* (the spoon with which I have eaten), here *cammac* cannot be relativized like **khAyA huA cammac* but in Telugu it is possible like *nenu tinna camcA* (The spoon which I used). When such situation occurs how do we arrive at from the source structure (Hindi) to target structure (Telugu)? Here comes the concept of *Transfer Grammar*, for which certain heuristics are required to map the linguistic or non-linguistic elements. Hindi has two types of relative clauses which are Correlative and Participle relative clauses (Yamuna Kachru, 1980). Correlatives are more common than participles. Furthermore, Correlatives can be divided into restrictive clause and non restrictive clause. Correlative clauses can be identified as they have explicit markers on them like jo....vo, jiskA...ve/vo/pronoun, whereas participle constructions have no explicit markers. On the other side,

---

[1]   The corpus that is used for the present study is not a parallel corpus.

[2]   AH is defined Subject–>Direct object–>Non-direct object–>Possessive –>Object comparison where '>' means more accessible than others

Telugu has two types of relative clauses (B.H. Krishnamurthy and J.P.L Gwynn, 1985; Usha Rani. A, 1990), they are *correlatives (clausal)* and *participles*, of the two types *participle* is a major type whereas *correlative* is used in the formal context. In fact, correlatives are also not that frequent as participles.

2 Relative clauses

One of the most important subordinate clauses is the relative clause. It is defined in literature as *a clause consists of necessarily of a head and a restricting clause or is a clause which qualifies or restricts the meaning of the noun in a noun phrase[3]*.

2.1 Different strategies of forming relative clause are
1. Relatives clauses can be Pre-nominal, Post-nominal and internal-headed relative clauses.
2. Between relative clauses where the role of the head noun in the relative clause is encoded +case versus -case.
3. Every language can relativize on subjects.
4. Any relative clause strategy must cover a continuous segment of the accessibility hierarchy (Keenan Edward. L and Bernard Comrie, 1977).

Basically, relative clauses are of two types in terms of semantics.
1. Restrictive relative clause
2. Non-restrictive clause (also called as appositive, descriptive, explanatory)

Both types might have similar syntactic constructions, but they are radically different in semantic/ pragmatic terms, particularly in that the restrictive clause uses presupposed information to identify the referent of a noun phrase, whereas the non-restrictive clause is a way of conveying new information on the basis of the assumption that the referent can already be identified. In typological terms, however this distinction seems to be almost completely irrelevant (Keenan Edward. L and Bernard Comrie, 1977).

2.2 Restrictive relative clause
These types of clauses serve to narrow down the entity designated by the modified noun phrase, in other words to help the hearer/reader to identify the referent of the common noun.

The young linguist whom I saw in the conference lives in Hyderabad.

Here the clause *whom I saw yesterday*, is a sub-ordinate clause, *whom* is the relative (wh) pronoun that serves to delimit the potential referents of the young linguist. The speaker assumes that the main clause does not provide the hearer with sufficient information to identify which young linguist he is talking about? Hence the (relative) clause is used.

2.3 Non-restrictive clause
These types of clauses serve merely to give the hearer an added piece of information about an already identified entity, but not to identify that entity.

The young linguist, who arrived for the conference is living in Hyderabad.

Here it is assumed by the speaker that *the young linguist* being talked about can identify by the hearer of language.

Language	Corpus size	correlatives	participles
Hindi	600	91	25

---

[3]

Table 1: Hindi relative clauses statistics from the corpus.

2.4 Difference

The difference between a restrictive and a non-restrictive interpretation is often only expressed intonationally. However, some languages including English and Hindi, use the comma (,) to indicate the non-restrictive clauses. Moreover, restrictive clauses do allow to drop the relative pronouns as

The young linguist I saw in the conference lives in Hyderabad.

But non-restrictive clauses do not allow it. Another distinction between a restrictive and a non-restrictive clause is the use of proper names in non-restrictive clauses as their heads, which restrictive clauses do not permit. For example *Chomsky, who arrived for the conference lives in Cambridge, Massachusetts.*

Indeed the general principle of relative clause formation is that relative pronoun is in the initial position, or at least part of a noun or a prepositional phrase. The accessibility hierarchy is defined by (Keenan Edward. L and Bernard Comrie, 1977), as Subject>Direct object>non-direct object>Possessive>Object comparison. It defines the ease of relative clause formation, i.e. it has an intuitive sense, easier to relativize Subjects, then Direct Objects then non-direct objects then that of Possessors than it is to Object comparison.

Clearly, however this intuition requires much more substantiation before it can be tested out rigorously.

3 Hindi Relative clauses overview

Hindi has two types of relative clauses, namely *correlative* and *participle clauses*.

3.1 Correlative clauses

I have used a corpus that consists of 600 sentences in which there are about 91 correlative relative clauses and 25 participle relative clauses are found (see Table 1). Table 1 indicates that correlative clauses are more frequent than *participles* in Hindi, which are the two types restrictive and non-restrictive. The structural features of relative clauses are,

1. The relativized noun is preceded by the relative marker jo (or jisase, jiake etc.),
2. The entire subordinate clause occurs in the sentence initial position[4],
3. The correlative marker vo (veh, ve etc.) occurs in the main,
4. And head noun has zero realization.

The following are some of the correlative markers. *jo-ve, jo-pronoun, jisake-ve/vo/pronoun, jisaseve/vo/pronoun, isako-/ve/vo/pronoun, jisaka-ve/vo/pronoun, jis-ve/vo/pronoun, jisame-ve/vo/pronoun, jahave/ vo/pronoun*. Depending on which NP is being modified, like subject, object, place, etc.

3.2 Restrictive relative clauses

a. *jo billi kA baccA mere gara mem hai vaha cUhe se DartA hai*

The kitten in my house is afraid of Rat.

Here *jo billi kA baccA mere gara mem hai* is a relative clause that occurs sentence initially. The marker *jo* the relative pronoun precedes the relativized noun *billi,* the correlative marker *vaha* occurs in the main clause *vaha cUhe se DartA hai,* and the head noun has zero realization in the main clause. The word order of the relative clause and the main clause is not fixed. There are many

---

4   But it can also occur in the non-initial position.

variations in the constituent order, depending upon thematization and focusing. The constituent of main clause and relative clause can be reordered (Yamuna Kachru, 1980).

- *vaha billi kA baccA cUhe se DartI hai* **[jo mere gara mem hai]**
- *vaha billi kA baccA* **[jo mere gara mem hai]** *cUhe se DartI hai*
- *cUhe se DartI hai vaha billi kA baccA* **[jo mere gara mem hai]**
- *cUhe se DartI hai vaha billi kA baccA* **[mere gara mem jo hai]**
- *hai vaha billi kA baccA cUhe se DartI* **[jo mere gara mem hai]**

However, it is not obligatory to have a zero realization of either the relativized or head noun.

> *jo billi kA baccA mere gara mem hai vaha billI kA baccA cUhe se DartI hai.*

Another sub-type within the restrictive clause where the post - head relative occurs when the head noun is indefinite (Yamuna Kachru, 2005; Yamuna Kachru, 1980) is give below.

> *eka student jo Apke sAt sAlse P.hd kar rahA hai bAhara baitA hai.*
> The student who is doing PhD with you for the last eight years is sitting outside.

Here the head noun occurs with an indefinite determiner *eka* (one) and the relativized noun following the relative marker *jo* has a zero realization.

3.3 Non-restrictive clause

The non-restrictive or appositive clause, which provides additional information about the head noun, follows the antecedent (see Figure 1).

Chandrababu Naidu, jo Andhra Pradesh kA mukhya
Chandrababu Naidu.NOM   REL. Andhrapradesh   Gen.M chief.M
mantrI thA Ajkal   yahAM   hai.
minister.M   was.sg   now a days   here   is

Chandrababu Naidu, who was the CM of AP is hear now a days.

Figure 1: Non-Restrictive Relative clause.

Similar case with the following sentence;

> *Mere bahan, jo hydarabad me hai, linguist hai.*
> My sister , who is in Hyderabad, is a linguist.

In both the cases the relative clause (*jo AP kA mukhya mantrI thA and jo hydarabad me hai*) is being used to give an additional information about an already identified entity. They have also been indicated with a comma. In correlative subject or object anything can be relativized. The above examples are for the subject relative. There are also other examples for direct object, indirect object, etc.

Direct object relativization

> *jo AdmI hame kal milA thA vaha dinnar par AyA.*
> The man whom we met yesterday came to dinner.

Indirect object relativization

> *jis AdmI ne tumhe pustak dI thI vaha dinnar ke liye AyA.*
> The man whom we met yesterday came to dinner.

Oblique relativization

It includes instrumental and locative instances. Locative relativization is as follows;

> *jo kurci pe mai baitA vo gir gayA.*
> The chair in which I sat fell down.

3.4 Participle relativization

In participle clauses there are no explicit markers. Hindi has two types of participles, known as *present participle and past participle* that modify the head noun as relative clauses do. Both types have no explicit markers. They both modify the noun. Both types inflect for the person, number and gender and function as either adjectival or adverbial clause. The present participle appears in the form of *verb-ta huA*, whereas the past participle appears in the form of *verb-A huA*. One distinction between both the types is that only a set of verbs yield past participle forms which indicate achievement, whereas all the other verbs have the present participle forms. In Hindi the accessibility hierarchy is lower than Telugu as Hindi allows only the subject (karta), and an object (karma) to make the participle construction other than that leads to ungrammaticality and unacceptablity.

Subject
　　dauDtA huA laDkA acAnak ruk gayA.
　　The boy who was running suddenly came to stop.

Direct object
　　rAmke dwArA likAgayA lek jarUr paRnA.
　　The letter which was written by Ram should be read.

Indirect object
As I mentioned earlier that participles cannot relativise indirect objects in Hindi.
　　*mere dwArA pustak diyA huA rAm gara calA gayA.

Similarly for locatives and instrumentals.
　　mere dwArA baitI hue kursI bahut mangi hui (locative),
　　*mere dwArA khAyA gayA cammac chandI hai (instruments).

4 Telugu Relative clauses overview

In Telugu also there are two types of relative clauses, namely, *correlative clauses* and *participle clauses*.

The present study, in fact, strongly supports as pointed out in (B.H. Krishnamurthy and J.P.L Gwynn, 1985) that the participles are more frequent than correlatives (see Table 3). The corpus has 600 sentences in which there are about 12 correlative clauses and 104 participle relative clauses are found (see Table 2). As for as Telugu relative clauses are concerned, they precede its head and also has the subject in nominative/accusative case like Hindi. Telugu also behaves like Turkish where subordination in general is by the means of non-finite construction (Chekuri Ramarao, 1969; B.H. Krishnamurthy and J.P.L Gwynn, 1985).

Language	Corpus size	correlatives	participles
Telugu	600	12	102

Table 2: Telugu relative clauses statistics from the corpus.

4.1 Correlative clauses

Correlative clauses which are found in Hindi are not normal in Telugu and the other Dravidian languages.

Hence, there are no correlative pronouns in Telugu similar to Hindi *jo-vo jab-tab*. However, with the influence of Sanskrit such syntactic constructions are slowly being introduced into Telugu and are used in formal speech (B.H. Krishnamurthy and J.P.L Gwynn, 1985). This is accomplished by using a question word in the subordinate clause + the complementizer [oo], followed by the main clause beginning with the corresponding demonstrative pronoun.

> *E ammAyi naccite A ammAyini pellicesuko.*
> Whom ever you like marry her.
> *eppudu Akalaite appude annam tinAli.*
> One must eat only when one gets hungry.

Constructions of the above types are more naturally expressed by relative clauses with participles.

> *nIku naccina ammAyini pelli cesuko, Akalainappudu tinu* respectively.

4.2 Participle relative clause

This type does not have any explicit relative morpheme. The verb takes on its attributive form and the modified noun becomes the head of the construction. A simple sentence can be changed into a relative clause by replacing its finite verb with a relative participle (or verbal adjective) in the corresponding tense-mode and shifting the noun that it qualifies as the head of the construction (B.H. Krishnamurthy and J.P.L Gwynn, 1985). The whole construction with the noun head becomes the noun-phrase in the matrix sentence as its subject, direct object of the finite verb, predicate complement or as an adverbial of time/place. The noun head of the relativized clause is deleted before a similar noun head in the matrix sentence. The relativized clause sIta pADina pATa derived from (1) is embedded before the direct object pATa of (2) as its modifier; the identical noun head pATa of the subordinate clause is then deleted, producing 3.

> 1. Sita pATa pADiMxi. 'Sita sang a song'.
> 2. Ramu pATa vinnADu. 'Ram heard the song'.
> 3. rAmu sIta pAdina pAta vinnAdu.
>   Ram heard the song which sita sang.

Now, we see which NP will be relativized.

Subject relativization

> *Adutunna abbAyi bAvunnAdu.*
> The boy, who is playing, is good.

Direct object relativization

> *rAmudu rAsina uttaram khaccitamgA cadavAli.*
> The letter, which Ram wrote must be read.

Hindi	Telugu
srujana	srujana
kAnA	annam
kA rahA hai	tin tunna di

Table 3: Lexical substitution.

Indirect object relativization

> *nenu pustakam iccina rAmuniki jvaram vaccindi.*
> Ram, whom I gave a book to, has got fever.

Genitive relativization

> *nenu cUsina rAmuni illu kUlipoyindi.*

134

Ram's house which I saw is destroyed.

Object of comparison
sarvAni vesukonna gounu kante jAhnavi vesukunna cokkA bAvundi.
The dress that Jahnavi bought is costlier than Bindu's dress.

4.3 Constraints

Though Telugu has the strategy of having rich participle relativization it has certain constraints on participles (Usha Rani. A, 1990). Noun with ablative case sometimes seems to be odd, and with directive post-positions like mIda (up), bayata (out) cannot be moved as heads of relatives with subsumed meaning.

5 Transfer Grammar

1. Transfer Grammar is one of the approaches to MT
2. Captures structural differences between SL and TL
3. The aim of the transfer grammar is to develop the translation system

It contains majorly two components, which are *lexical transfer* and *structural transfer*. Since both languages have similar syntactic word order, it does not require any structural transfer for this pair of sentences (see table 3).

5.1 Relative clause Rules

Let's look at the correlatives in Figure 2. If we map (Figure 2 to Figure 3) between main clause to main clause (*vah bahut nuksAn kar gayI - cAlA nasTam cesindi*), there seems to be no problem except the correlative marker *vah*. Now let's take the subordinate clause mapping. There are two differences. The order of the TL differs from the SL and SL has finite form *AI thI* in the clause whereas the TL has non-finite form *vaccina*. The differences between SL and TL are as follows:

1. SL has explicit relative markers, whereas the TL has no explicit markers
2. On the one hand both clauses are finite in SL, on other it is not, the relative clause is non-finite
3. SL and TL word order is different (see figure 3)

How can this be mapped? Here comes the concept of *Transfer grammar* to bridge the gap where the differences are mapped to get the expected structure. One would form the following rules to map the SL structure to the TL structure.

1. Change the word order according to the TL
2. Delete the Relative markers (jo and vah)
3. Convert the finite verb into Non-finite from the jo (sub-ordinate) clause

jo	ANdhI	kal	AyI		thI	vah	bahut
Rel.	storm.F	yesterday	come.Pef.F.sg		Past.F.sg	that	much

nuksAn	kar	gayI.
damage.M	do	go.PerfF.sg

ninna	vaccina	tufAnu	cAlA	nasTam	cesindi.
Yesterday	come.past.verbaladj	storm.Nom	much	damage	did.3.sg.Neu

'The storm that raged yesterday did a great deal of damage'

Figure 2: Hindi and corresponding Telugu relative clause.
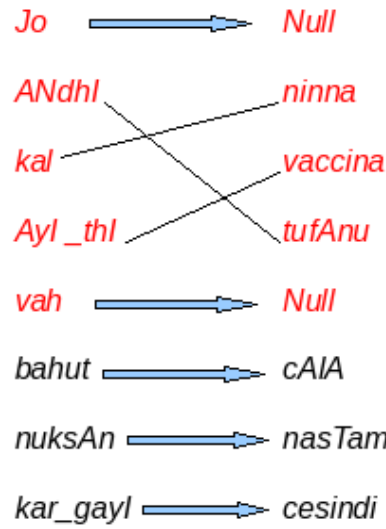
135

Figure 3: Hindi and Telugu clause Mapping.

After applying the rules respectively we get the perfect translation of the sentence (see figure 4).
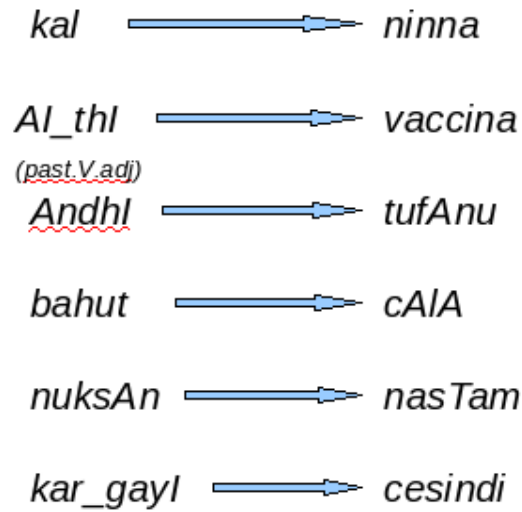

Figure 4: After applying rules.

6 Conclusion

In this paper, the corpus based study of relative clauses; distribution and their nature in both Hindi and Telugu are discussed. The paper also talked about transfer grammar and its significance. Then it discussed various issues involved in mapping relative clauses from TL to SL from Machine Translation perspective.

However, it should be noticed that for doing a better Machine Translation, apart from transfer grammar, other resources are required like broad coverage morph-analyzer, generator, machine

readable bilingual lexicons, large amount of POS and parsed tree corpus and large amount of parallel corpora of the source and target languages.

References

.Krishnamurthy ,B.H. 1985.*A grammar of modern Telugu*. Oxford University Press, Delhi..

Krishnamurthy ,B.H and P Sivenanda sharma. 1968b.*A basic course in modern Telugu*. Osmania University, Reprinted by Telugu academy (2001), Hyderabad.

Ramarao. ,C 1969.*A grammatical sketch of Telugu*. Commissioned by the Ministry of Home affairs, Language Division.

C Ramarao. 1982.*Telugu lo velugulu*. Hydearabad Andhra Saraswata Parishattu.

Keenan Edward. L and Bernard Comrie. *Noun phrase Accessibility and Universal Grammar*. Linguistic inquiry. 1977.

McGregor R.S. 1972..*Outline of Hindi Grammar, 3rd edition*. Oxford niversity Press.

Agnihotri ,Rama Kant. 2007. *Hindi, an essential grammar*. Routledge Taylor and Francis group. Rebecca Root. 1985.

…. *A two way approach to structural grammar in MT*. In Proc. EACL, pages 70-72, Geneva, Switzerland.

Rani ,Usha 1990.A. *Relativization in Telugu*. OPIL, Osmania University.

Kachru. ,Yamuna 1980.*Aspects of Hindi grammar*. New Delhi: Manohar.

Kachru. ,Yamuna 1980.*A Contemporary grammar of Hindi*. Macmillan Company of India.

Kachru. ,Yamuna 1973.*Relative clause formation in Hindi*. Basha, Delhi.

Ecolinguistics, Semantics and Pragmatics – Two Case Studies: The *Hit Chaurasi Pad* and Norfolk Island, South Pacific

Joshua Nash
University of Adelaide,
Australia
joshua.nash@adelaide.edu.au

Abstract

Ecolinguistics or the scientific study of language ecologies and the interaction between language and the environment began as a subdiscipline of applied linguistics in Western universities in the late 1970s. It is primarily concerned with:

- How the contours of the natural environment match linguistic contours in respective milieus and vice versa – the ecology of language, and
- Environmental discourse analysis – the language of ecology.

Despite several seminal works and although much has been written in various volumes covering the theory and epistemology of ecolinguistics, the field still awaits further theoretical expansion and synthesis within itself and together with other fields of enquiry. The multidisciplinary focus of ecolinguistics requires that ecolinguists take a holistic approach embracing a multitude of perspectives both Eastern and Western. It is proposed here that the discipline should strive ahead utilising modern linguistic and environmental theories and technologies while maintaining the validity of ancient basics and practical measures that are the foundation of the area of enquiry.

This paper proposes a synthesis of ecolinguistics with:

1) An Indian approach to the treatment of language and environment, i.e. the semantic aspect, and
2) The case study of language evolution and change on Norfolk Island, South Pacific, i.e. an example of communicative adaptation – the pragmatic aspect.

It employs two facets for analysis:

1) The *Hit Caurasi Pad* (*Eighty-four Stanzas*) of medieval Indian poet Hit Hari Vansh Goswami for its linguistic descriptive power and artistic depiction of humans interacting with Nature in a divine way, and
2) Place names on Norfolk Island in the Norf'k language for their practical, historical and environmental implications and significance.

These two case studies aim at finding some suggestions to advancing current thought and theory in ecolinguistics through:

1) Considering the importance of Indian/Eastern aspects of human-environment interaction; and
2) Suggesting the advantage of doing ecolinguistics and placename study on small island environments to hint at various universals and commonalities in linguistic and ecological contact.

1 Introduction

The relatively new and multifaceted discipline of ecolinguistics, environmental linguistics or language ecology has been explicated in various volumes (Harré et al. 1999; Fill 1996) and recently from a dialectical perspective (Bang & Døør 2007). The field of enquiry can be broadly divided into two parts:

1. The scientific study of how the contours of language match the contours of the natural environment, and

2. Environmental discourse analysis.

These two aspects have been presented in Fill and Mühlhäusler (2001), further developed by Mühlhäusler (2003) in the first coursebook in ecolinguistics and from a critical discourse perspective by Alexander (2009). Apart from an apparent increased interest in studying the relationships between language and environment with the appearance of the online journal *Language & Ecology* (www.ecoling.net/journal.html) and a wikipedia.org site for ecolinguistics (en.wikipedia.org/wiki/Ecolinguistics), there has not been the warranted theoretical push to solidify the discipline on the international linguistic and environmental stage. This was noticeable in Germany last year at the 15th World Congress of Applied Linguistics (AILA) where only around 20 linguists presented papers relevant to ecolinguistics. Of these papers all were given by academicians from European and Australian universities, hence the relevance and applicability of the discipline outside these circles is raised. The first AILA session in ecolinguistics was held in 1990 and since that time there has not been a great increase in the absolute number of ecolinguists attending such sessions. Furthermore the theoretical weight of these presentations for taking the discipline further is questionable as many of these papers merely reinterpreted old ideas and were limited both in terms of empirical analysis and theoretical scope. What I hope to open up here is a new way of looking at ecolinguistics in light of an Indian/Eastern perspective with the aim to:

1. Take ecolinguistics beyond its strong Western foundations and analyses, and

2. Offer stronger theoretical and practical considerations for ecolinguistics, which following from the above seems to be needed.

Here I attempt to move away from the current trends in, for example, environmental discourse analysis and move towards an initial argument and basis for which the broader theoretical points of the future of ecolinguistics can stand. In this paper I will only be presenting a small subset of the data that I have amassed on field trips to Norfolk Island. It is this larger data set and evolving theoretical framework that I will use to expand the conceptual foundations of ecolinguistics. This is definitely not the last word on this matter and it is quite an ambitious undertaking to employ medieval Indian poetry in parallel with an analysis of placenames on a small island in the South Pacific Ocean to speculate, consider and arrive at a practical theory, philosophy and understanding of language, nature, place and life. This said, ecolinguistics has always strived for a synthesis in understanding the human being and its mental, cognitive and linguistic relationships in the environment and the world so being game and proposing a new approach is certainly welcomed.

2 Ecolinguistics: From Broad to Focussed

Surveying the literature in ecolinguistics reveals that although there has been a fair amount of theoretical development, the empirical focus and use of detailed case studies to illustrate theoretical findings and proposals has certainly gone wanting. Mühlhäusler and Peace (2006) have summarised much work in ecological discourse and give a good general survey of the discipline while Alexander (2009) presents a more empirical study of many of the principles important to analysing the linguistic and philosophical aspects of environmentalism and their discussion in the media. These approaches to environmental discourse analysis offer several keys to open up texts to linguistic analysis by:

1. Looking in detail at lexicogrammatical facets of texts, and

2. Drawing conclusions regarding the contextualisation of texts.

The empirical approach to ecolinguistics has mainly been limited to:

1. Analysis of word lists and folk lexicon, e.g. Schultz (2001), Little (1999),

2.Grammatical analyses, e.g. Goatly (2001), Halliday (2001), and

3.Environmental discourse analysis, e.g. Alexander (2009), Mühlhäusler (2001).

My argument and criticism of research hitherto conducted in ecolinguistics is that it has not focussed on case studies where these linguistic levels are applied to real data eliciting palpable, usable and replicable results and that where empirical data has been analysed, the broader applicability of results is questionable as the conceptual foundation of the case studies is too broad. This is where I propose the treatment of two extremely focussed data sets – a small list of local, esoteric placenames on Norfolk Island, South Pacific, and verse 12 of Hit Hari Vansh Goswami's poem the *Hit Chaurasi Pad* – for their power to propose some theoretical direction in and for ecolinguistics. Further I aim to argue that by analysing two data sets on either extreme of the pragmatic (communicative and informative function, i.e. placenames on Norfolk Island) and semantic (aesthetic, essential and meaning-based qualities, i.e. *Hit Chaurasi Pad*) continuum suggestions toward a model of the relation between language in practice and language in essence will arise. This will in turn put forward a framework for understanding the interaction between language and nature or *bhasha* and *prakriti* in the Indian context respectively.

3 Research Rationale

The data sets and locations, Norfolk Island and Vrindavan, were selected for practical purposes: I have experience and have been involved in longitudinal research into observing relationships between language, culture and environment for several years in both locations. In this sense I treat Norfolk Island as an ecolinguistic snapshot, i.e. a manageable case study for observing relationships between language, human perception of the environment and environmental management, and specifically the erotic poetry of the saints of Vrindavan as important philosophical underpinnings for the modern environmental movement that arose in the 1970s dedicated to the conservation of the medieval holy town that was once a verdant forest and playpen for the enjoyment of the Divine Couple and representatives of the Cosmic Duality, Radha-Krishna. Location maps for both settings are given below in Figure 1 and Figure 2:
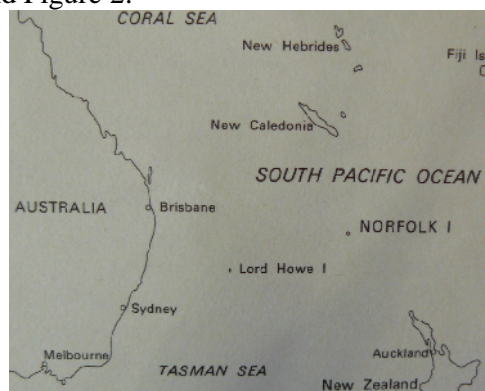

Figure 1 - Norfolk Island location map

Figure 2 – Vrindavan location map

My main assumption in undertaking this analysis is that by using two minimal data sets and exposing them to a research framework common in ecolinguistics concrete results proposing some basic principles about how human beings speak about and orientate themselves in the environment will arise.

4 Methodology

My argument is based in observing interactions between lexicon and its ability to describe the environment and events that took place in this environment of a small data set of five unofficial Norfolk Island toponyms, i.e. placenames that do not appear on official maps, which I have acquired through archival and interview research on Norfolk Island during fieldwork in 2007 and 2008. This data is provided below in Table 1. These placenames are described and considered with reference to their placement within the folk lexicon of Norf'k, the language of Norfolk Island, and how these esoteric placenames illustrate how users of these names conceptualise their natural and cultural environment and orientate themselves in history and place and time and space. Here the pragmatic domain of communication is emphasised.

Hari Vansh's canonical work, the *Hit Chaurasi Pad*, has been analysed from several historical and hagiographical perspectives (see particularly Snell 1991). My methodology though observing an entirely different genre of data is essentially the same as the method employed to evaluate Norf'k placenames: how do humans cognise the world and orientate themselves linguistically in nature and moreover how does a poet illustrate and make famous the ability of his characters through metaphor and imagery in seemingly endless time and space to be in consonance with real places and real environments that still exist today in the modern day city of Vrindavan, India? The semantic, artistic and essentialist domain of communication here is emphasised. What I offer is initial research into a proposed new translation and interpretation of the *Hit Chaurasi Pad* in English from an environmental angle as well as suggesting the relevance of the work to broad ecolinguistic analysis.

In conclusion I compare the results of the respective analyses and relate the combined outcomes of the research to ecolinguistic theory while proposing future and further directions for the field as a whole.

5 Norfolk Island, South Pacific: An Ecolinguistic Snapshot

Elsewhere I have outlined the relationship between Norfolk Island and the Indian context (Nash in press). On the surface there may not appear to be many parallels between the analysis of placenames in a South Pacific contact language and a medieval devotional poem in Braj Bhasha, the dialect of Hindi spoken in the Braj area of northern India, but here I propose that it is these two examples each on either extreme of the semantic-pragmatic continuum that will bear some ecolinguistic fruit.

The Norf'k language is a contact or mixed language spoken by approximately 300 descendents from Pitcairn Island on Norfolk Island, a small island of 36 square kilometres in size 1700 kilometres east of Sydney and situated between New Caledonia and New Zealand. The status of the language has received a fair amount of research interest while the typology of the language remains unclear (Mühlhäusler 2007). My concern here will be with an aspect of research into Norf'k that remains largely unwritten, viz., Norf'k placenames and more specifically unofficial, esoteric insider placenames. Norfolk Island's starkly demarcated historical periods provide historians and linguists with relatively clear-cut historical periods and reliable and traceable data concerning the human-environment situation at particular points in time.

A very deliberate list of five Norf'k placenames was chosen for its descriptive power and for its ability to convey meaning-laden insider sociocultural, historical and environmental concepts that are so prevalent in the Norf'k language. These names were all acquired during interview based fieldwork with informants on Norfolk Island in February 2008. The list is presented below in Table 1:

Norf'k	English	Etymology
1.Johnny and Nigger Bun Et	Johnny and Nigger Burnt It	3.An area where two local men lit a fire that burnt out of control.
2.Side Ar Whale Es	Literally 'the place where the whale is' or 'Whale('s) Place)	A land feature which when looked at from a distance resembles a whale.
3.Parloo Park	Masturbation Park	An area young boys and girls used to frequent on their first meeting.
4.Foote Nort	Why not	Eldon Foote's property; humorous and quirky allusion to common Norf'k expression 'foot nort' English: why not.
5.Em Steps	The Steps	The Norf'k name for the convict steps on Nepean Island, the small island 800 metres south of Norfolk Island; also known as 'Dem Steps'.

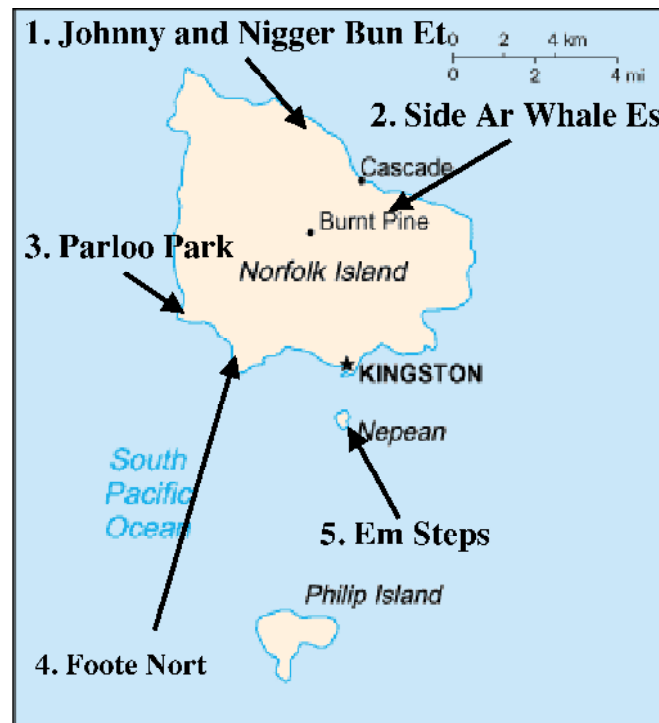The location of these places on Norfolk Island is presented diagrammatically below in Figure 3:

Figure 3 – Location of five Norf'k placenames on Norfolk Island

Place names are linguistic, cultural and environmental markers of history and tools to create, perpetuate and narrate stories. This situation is more pronounced when:

1. There is language contact where conflicting and complimentary stories between and across cultures can arise, and
2. A group of people is forced to speak about, manage and live in a new and unfamiliar environment.

Example 1 illustrates among other things a common trait of Norf'k phonology namely word final consonant loss, i.e. Norf'k 'bun' for English 'burnt'. This is possibly a residual Polynesian feature in the language (cf. Ross & Moverley 1964) reflecting its linguistic connection to Polynesia and Tahiti and the initial contact language that developed on Pitcairn Island after the Bounty mutiny took place in 1789. 'Johnny and Nigger Bun Et' further refers to two well-known Norfolk Islanders and the activity that led to the naming of the area, lighting wood fires so the whalers offshore would not lose their way back to shore at night, solidifies an element of Norfolk's cultural landscape into a name associated directly with a place on the north coast of the island few people know about.

Example 2 provides evidence of the descriptive power of the landscape and how humans name places based on their perception of a topographical feature and possibly its utilitarian nature. One informant intimated to me that 'Side Ar Whale Es' looks like a whale if one looks at it in the right manner and uses a bit of imagination. It also shows how esoteric names can be as this name was only known by one of my informants and it was most likely he who named it. This name also illustrates various distinguishing aspects of Norf'k syntax from English – the formation of the genitive construction, e.g. 'Whale('s) Place' is one possible translation of 'Side Ar Whale Es' in English. The work Norf'k syntax does in and by producing such constructions in placename grammar as well as in spoken sentences demonstrates the power the language possesses in its potential to demarcate itself from the linguistic stranglehold of English. Here we see the strength of

Norf'k in creating sociolinguistic boundaries and how it forges strong relationships with people and place, namely the folk and people of Norfolk Island.

'Parloo Park', literally 'Masturbation Park', is an area in the south west of Norfolk which has been known to be a haunt for young couples' courting behaviour. That locals remember and have named this locale indicates a strong event based memory of the happenings in the area and reflects the phenomenon of action and incident in relation to place being crystallised into language use. The name also reveals the presence of Polynesian elements in the language and further the use of Tahitian lexicon for taboo elements in (particularly) human behaviour and bodily functions once again depicts a strong link between individuals and language in the Norfolk setting. This is also one of the few Norf'k placenames of distinct Polynesian origin.

Example 4 'Foote Nort' presents a placename directly associated with a person, Eldon Foot, who used to live in the Rocky Point area. The humorous and quirky allusion to the common Norf'k expression, 'foot nort' (English: why not), is based on phonological similarity between the Norf'k adverb and the gentleman's name. This exemplifies an aspect of unofficial naming that is often overlooked – people name places for fun and to pass time. In this example Eldon Foote, a Canadian philanthropist, has been linguistically concretised in place and space in the Norfolk environment while at the same time lexified into the folk lexicon and intimately associated with the Norf'k language, although Foote was not a Norfolk Islander or Norf'k speaker.

'Em Steps' and the variation 'Dem Steps' (English: The Steps) refers to the convict steps on the eastern coast of Nepean Island, a small uninhabited island part of the Norfolk archipelago 800 metres south of Norfolk. The steps and the subsequent English name, 'The Convict Steps', developed during sandstone excavation that took place throughout the convict settlement on Norfolk prior to the arrival of the Pitcairners in 1856. Although there is no direct relation between the Norfolk Islanders and the events that created the convict steps as a topographical feature, a Norf'k name has developed and is an integral part of the folk lexicon and toponymy of Nepean Island, especially among the local fishers. This additionally emphasises the difficulty in the semantics of delineating what constitutes a Norf'k or an English name in the complicated tapestry of language contact in the toponymy of Norfolk Island. That is, syntactic and phonological analysis of names will often not give us clues as to how Norfolk Island toponyms should be classified. These five examples and the theoretical development arising from this brief analysis will be discussed separately and in relation to the findings relevant to ecolinguistics from the *Hit Chaurasi Pad* in section 7 below.

6 Vrindavan: The Terrestrial Stage

The cultural and social aspects of Vrindavan have been explicated by Growse (1882) in the first detailed expose of the Radha-Vallabh temple tradition to the English speaking world. It is this tradition of the worship of Radha-Krishna and the founder Hit Hari Vansh Mahaprabhu's seminal poem *Hit Chaurasi Pad* (*Eighty-four Stanzas*) that form the philosophical basis of this analysis.

I have chosen one complete stanza of the *Hit Chaurasi Pad* in two different translations to look at in detail for its environmental and linguistic aspects. Verse 12 of the *Hit Chaurasi Pad* was selected for pragmatic reasons – it is the first verse in the work that explicitly:

1. States precise geographical locations in Vrindavan, e.g. Yamuna, Vamshivat,
2. Employs complex environmental imagery and relies on the use of descriptive metaphor, e.g. 'where the earth is supremely beautiful' and 'moonlight is brilliantly clear' (White 1977: 61), and
3. Refers directly to the flora of Vrindavan, e.g. 'new bower' and 'jasmine is budding slightly' (Snell 1991: 193)

144

and synthesises them into an analysable whole. I now employ Verse 12 of Hari Vansh's aesthetic treatise on nature, love, life and spirit to unpack and develop the locational, environmental, metaphorical and botanical parameters of a new theoretical focus in ecolinguistics. The two translations of Verse 12 of the *Hit Chaurasi Pad* are presented below:

1   Come, wise Radhika! For your sake Shyam has arranged a round-
      dance,
      a store of joy, on the bank of the Yamuna:
2   groups of young girls dance in great eagerness at the music and merriment
              as the joyful flute, source of delight, is playing.
3   In that most pleasing place near the *vamshivata*
              a soft breeze blows from the [sandal-clad] Malaya mountain,
              yielding all joys.
4   the forest is strongly fragrant with half-blown jasmine,
              and there is bright moonlight in the full-moon autumn night.
5   Cowherd girl, feast your eyes on Naravahana's Lord,
              whose head-to-toe beauty removes the agony of desire;
6   lady! Experience this ocean of delight, rejoice with your arms
              joined around his neck.
For Shyama's sport in the fresh bower is worthy of the world's praise!
  (Snell 1991: 193)

TWELVE
O clever Radha, come along!
For Your sake, Shyam, the Abode of Pleasure, has brought forth
The Rasa Lila on the banks of the Daughter of Kalimda (Yamuna).(1)

A company of maidens dances in front of him:
They are filled with great joy at the raga's tone.
He plays his blessed flute, the source of rasa. (2)

Around the Vamshivat (the tree beneath which Krishna stands to entice the gopis) there,
Where the earth is supremely beautiful,
The scent of sandalwood blows on the gentle air, giving ease to all. (3)

The jasmine is budding slightly:
The forest is extremely fragrant. On the full-moon night
Of the month of Sharad the moonlight is brilliantly clear. (4)

The Lord, the Vehicle of Salvation for men,
Gazed and filled his eyes with the Herdsman's Daughter (Radha) –
Beautiful from head to foot, the destroyer of amorous distress. (5)

O Lady, enjoy his arms encircling your neck.
Bear the Ocean of Pleasure!
Shyam's love play in the new bower is worthy of the praise of the world. (6)
(White 1977: 60-61)

I begin by first dividing aspects of the stanza into three categories: places, flora and environmental events/things. These are presented below in Table 1 (Snell translation) and Table 2 (White translation):

Places	Flora	Environmental events/things
(Bank of the) Yamuna (1) Vamshivata (2) Malaya mountain (3) Most pleasing place (3) Forest (4)	Half-blown jasmine (4) Fresh bower (6)	Soft breeze (3) Bright moonlight (4) Full-moon Autumn night (4) Ocean of delight (6)

Table 1 – Environmental Aspects of Stanza 12 of *Hit Chaurasi Pad* from Snell (1991: 193)

Places	Flora	Environmental events/things
On the banks of the Daughter of Kalimda (Yamuna) (1) Vamshivat (2) Forest (4)	Jasmine is budding slightly (4) New bower (6)	Where the earth is supremely beautiful (3) Jasmine is budding slightly (3) Forest is extremely fragrant Scent of sandalwood blows on the gentle air (3) Moonlight is brilliantly clear (3) Full-moon night (4) Ocean of pleasure (6)

Table 2 – Environmental Aspects of Stanza 12 of *Hit Chaurasi Pad* from White (1977: 60-61)

Hari Vansh locates his work in time (full-moon Autumn night) and space/place (Yamuna bank near Vamshivata). Situated in the north-eastern part of Vrindavan, Vamshivat is depicted linguistically as the celestial bower and playpen of Shyamaa-Shyam or Radha-Krishna. He introduces the scene and setting by relating the 'most pleasing place', Vamshivat on the Yamuna bank, to the natural surroundings - Malaya mountain, supremely beautiful earth, budding jasmine, soft breeze - and in turn introduces the human aspect in the latter part of the verse to this exquisite natural setting. It is the human in the form of Shyam (Krishna) constructed by an array of metaphors and how this depiction relates to the natural environment that is of interest to ecolinguistics. Metaphorically Krishna is constructed as:

1. The Vehicle of Salvation,
2. The Abode of Pleasure,
3. The Ocean of Pleasure, and
4. The Ocean of Delight.

Krishna, the provider of liberation, is likened to a thing – a vehicle, abode and ocean – fixed sturdily and with strength within the pleasure environment of the fresh bower created for cosmic interaction with his consort, Radha, 'the destroyer of amorous distress'. Humans are equated to special thought interactors in the lap of nature whose duty is to enjoy the cosmic love play in the bower and be active participants in the sport that "is worthy of the praise of the world" (White 1977: 61). Here I emphasise that language and specifically the semantic facet of communicating, e.g. word sense, metaphors and interpretations of meaning laden lexemes, all contribute to an environmental reading of Hari Vansh's work.

Words and concepts relating to environmental events/things, e.g. 'scent of sandalwood blows on the gentle air', earth, forest, 'bright moonlight', all represent a conception of nature, here symbolised and individualised by Hit Hari Vansh and translated from Braj Bhasha Hindi to English,

which epitomises the cognitive capacity and adaptability of humans to abstract things conceptually from the world, create names for things which are of use to them, relate these creatively and artistically to myths and stories passed down through the ages and live a life in close presence and consonance with nature and the rest of human society.

Humans, words and nature in the poet's view are all a part of the cosmic system and thus words and how we describe nature affect our perception of our natural environment; Vrindavan is an example of the dialectic between human and nature and language and thought (Bang & Døør 2007) and Hari Vansh's contribution helps us consider the relationships between sound and silence, word and sentence, meaning and function and subject and object. Hari Vansh's word illustration of a canvas of interaction between the human, mind, nature and cosmos presents a new realm and possibility for ecolinguistics to move beyond its firmly Western foundations and into an Eastern understanding of the conception of language and particularly semantics and invite and encourage further theoretical development from this differing perspective.

7 Discussion and Conclusion

Referring to the tasks I set myself at the beginning of this paper I have attempted to put an ecolinguistic analysis into an Eastern conception of language and nature and have coupled this with an extremely focussed study of very specific empirical and real data from what I claim is an appropriate case study for ecolinguistics, namely Norfolk Island. I emphasised the pragmatic domain of communication with respect to Norfolk as:

1. The insider placenames described form an important part of the descriptive spatial and orientational lexicon of a distinct section of the Norfolk population which is used in their everyday communication, and

2. Through usage these names in direct contact with the place and environment of Norfolk have become solidified into the Norf'k language.

The historical and environmental parameters of this section of the Norf'k lexicon lend themselves very well to adaptation and environmental change due to their unofficial and folk nature. It is certainly these facets and functions of language where traditional classification systems and analytical tools, e.g. conventional methods for distinguishing between and describing pidgins and creoles and other contact languages (see Sebba 1997, Reinecke et al. 1975), do not fit that must be at least one of the foci of further studies into the pragmatic arena of ecolinguistic theory. I suggest that small case studies based in interaction with real people, real data and vibrant communication situations provide an approach to exploring this element of language and environment interaction.

On the other end of the communication spectrum I have used Hari Vansh's *Hit Chaurasi Pad* to emphasise the relationship between semantics, sense and place and how these are communicated artistically in an English translation of a devotional poem. My claim, however, is that although these case studies are on opposite ends of the pragmatics – semantics continuum and illustrate different levels of abstraction from real communication situations, the principles involved in naming things, interacting with the environment cognitively and linguistically and creating meaning and attachment to space and place are essentially the same: there are utilitarian reasons, practical motives and situational and historical bases for names and the practice of naming things and ultimately how humans locate themselves in a given environment. Further reflection and analysis of available data will bring forth additional findings into the complex relationships that ecolinguistics seeks to explore. This I claim is where the steps forward in furthering the theoretical development and synthesis of ecolinguistics in the West and the East are required. It is with this sense of hope, anticipation and eagerness for continued research into understanding interactions between humans, language and environment that I conclude with an invitation to my readers and co-researchers.

References

Alexander, R.J. 2009, *Framing Discourse on the Environment: A critical discourse approach*, New York: Routledge.

Bang, J.C. & Døør, J. 2007, *Language, Ecology and Society: A dialectical approach*, S.V. Steffensen & J. Nash (eds.), London: Continuum.

Fill, A. (ed.) 1996, *Sprachökologie und Ökolinguistik.* Tübingen: Stauffenburg.

Fill, A. & Mühlhäusler, P. 2001, *The Ecolinguistics Reader*, London: Continuum.

Goatly, A. 2001, 'Green Grammar and Grammatical Metaphors or Language and the Myth of Power or Metaphors We Die By', in A. Fill & P. Mühlhäusler, 2001, *The Ecolinguistics Reader*, London: Continuum, 203-225.

Growse, F.S. 1882, *Mathura: A district memoir*, Allahabad: North-West Provinces & Oudh Government Press.

Halliday, M.A.K. 2001, 'New ways of meaning: the challenge to applied linguistics', in A. Fill & P. Mühlhäusler, 2001, *The Ecolinguistics Reader*, London: Continuum, 175-202.

Harré, R., Brockmeier, J. & Mühlhäusler, P. 1999, *Greenspeak: A study of environmental discourse*, London: Sage.

Little, P.E. 1999, 'Environments and environmentalisms in anthropological research: facing a new millennium, *Annual Review of Anthropology*, 28, 253-84.

Mühlhäusler, P. 2007, 'Why Pitkern-Norf'k is not a creole', Paper presented at Australian Linguistics Society conference September 26-28, 2007, Adelaide, Australia.

Mühlhäusler, P. 2003, *Language of Environment – Environment of Language: A course in ecolinguistics*, London: Battlebridge.

Mühlhäusler, P. 2001, 'Talking about environmental issues', in A. Fill & P. Mühlhäusler, 2001, *The Ecolinguistics Reader*, London: Continuum, 31-42.

Mühlhäusler, P. & Peace, A. 2006, 'Environmental Discourses', *Annual Review of Anthropology*, 35, 457-79.

Nash, J. in press, 'Comparing Linguistic Landscapes in Vrindavan, Uttar Pradesh and Norfolk Island, South Pacific', *Proceedings from Perspectives in Linguistics 2008, University of Kashmir*, Srinagar: Department of Linguistics, University of Srinagar.

Reinecke, J.E., Tsuzaki, S.M., De Camp, D.E., Hancock, I.F. & Wood, R.E. (eds.) 1975, *A Bibliography of Pidgin and Creole Languages*, Honolulu: University of Hawaii Press.

Ross, A.S.C. & Moverley, A.W. 1964, *The Pitcairnese Language*, London: Deutsch.

Schultz, B. 2001, 'Language and Environment', in A. Fill & P. Mühlhäusler, 2001, *The Ecolinguistics Reader*, London: Continuum, 109-14.

Sebba, M. 1997, *Contact Languages: Pidgins and Creoles.* London, Macmillan.

Snell, R. 1991, *The Eighty-Four Hymns of Hita Harivamsa: An edition of the Caurasi Pada by Rupert Snell*, New Delhi: Motilal Banarsidass.

White, C.S. 1977, *The Caurasi Pad of Sri Hit Harivams: Introduction, translation, notes, and edited Braj Bhasa Text*, Honolulu: University of Hawaii Press.

Revised receptive expressive emergent language scales for Kannada speaking children

Madhu.K	Ms. Deepa M.S.	Suhas K.	Harshan Kumar H.S.
madhuknayak@gmail.com	deepalibra@gmail.com	suhas.ck@yahoo.com	harshankumar@rocketmail.in

JSS institute of speech and hearing
Ooty road, Mysore- 570025

Abstract

Language development is a process that starts early in human life, when a person begins to acquire language by learning it as it is spoken and by mimicry. Child language development move from simplicity to complex. Many tests have been developed for language in toddlers. Even though they have been developed many decades back, they are still in practice in almost all clinics in India. But the tests need to be revised because children are observed to be developing many skills at a very early age.

1 Introduction

The Language development is a process that starts early in human life, when a person begins to acquire language by listening to spoken language and by mimicry. Children's language development moves from simplicity to complexity. Infants start without language. Yet, by four months of age, they can read lips and discriminate speech sounds.

Usually, language starts off as a recall of simple words without associated meaning, but as children grow in age, words acquire meaning and connections between words are formed. Over time, sentences start to form as words are joined together to create logical meaning. As a person grows older, new meanings and new associations are created and vocabulary increases as more words are learnt.

Infants use their bodies, cries and other preverbal vocalization to communicate their wants, needs and dispositions. Even though most children begin to vocalize and eventually verbalize at various ages and at different rates, they learn their first language without conscious instruction from parents or care takers. It is a seemingly effortless task that grows increasingly difficult with age. Of course, before the learning can begin, the child must be biologically and socially mature enough.

The most intensive period of speech and language development for humans is during the first three years of life, a period when the brain is developing and maturing. These skills appear to develop best in a world that is rich with sounds, sights, and consistent exposure to the speech and language of other individuals.

There is increasing evidence suggesting that there are "critical periods" for speech and language development in infants and young children. This means that the developing brain is best able to absorb language, any language, during this period. The ability to learn a language will be more difficult, and perhaps less efficient or effective, if these critical periods are allowed to pass without early exposure to a language. The beginning signs of communication occur during the first few days of life when an infant learns that a cry will bring food, comfort, and companionship. The newborn also begins to recognize important sounds in his or her environment. The sound of a parent or voice can be one important sound. As they grow, infants begin to sort out the speech sounds (phonemes) or build blocks that compose the words of their language. Research has shown that by six months of age, most children recognize the basic sounds in their native language.

Language development during these very early years of life from birth to about 2 ½ to 3 years is very dramatic and rapid. From birth to 2 ½ years is the period usually labeled as "Infancy". Infants eventually recognize much of what they hear and gain more control over their speech structures. Even though they may respond to hearing a familiar word or produce a string of sounds that are almost recognizable, their abilities are still pre-linguistic, that is, they precede true language. Nonetheless, they represent the needs of conventional language behaviors.

By the first year children start babbling and motor milestones also reach first step. During the second year children start speaking one word sentences and the vocabulary increases. As in other areas of linguistic research, it is important to recognize that different constraints act upon the child's comprehension and production of a particular form. Language does not develop in isolation as a separate system of behavior with special status. The behaviors that eventually evolve into recognizable language behavior are supported by the child's holistic development of the motor, cognitive, and social domains.

The domains (motor, cognitive, and social) are interdependent in their development process. Although this study focuses on language, it is important to view the entire process and to understand how these related domains set the stage for language to evolve.

In the 1st year many changes occur rapidly in all the domains. These changes continuously channel the infant's evaluation towards true language. Hence the area of interest was to measure the language in the typically developing toddlers. As we know that language develops rapidly during the critical period, it is important to know the domains of language that is acquired during this stage. Also the level of acquisition of language in infancy was studied. Many researchers have listed the skills that were achieved across different stages of language development. Many of these serve as tests to assess language abilities in children with language impairment.

RECEPTIVE EXPRESSIVE EMERGENT LANGUAGE SCALE (REELS) was given by Bzoch and League (1971) for children in the age range of 0 to 3 years. It is an untimed test although may take approximately 10 minutes. The test aimed at determining whether expressive and receptive language skills were following normal developmental patterns during first 36 months of life. The test consists of an outline of developmental stages for expressive and receptive language presented in 22 sections. There were 3 receptive and 3 expressive items per each section and these were divided across the age bands so that 12 items apply to year one, 6 items apply to year two, 4 items apply to year three.

Bzoch, League, and Virginia (2000), Revised the REELS (Receptive Expressive Emergent Language Scales) using 1112 children. It is designed to help clinicians to identify infants and toddlers who have language impairments or who have other disabilities that affect language development. It is especially useful as an assessment and planning instrument in Early Childhood Intervention programs.

The tests are subjected to revision because of many reasons.

Importantly the tests which were used to assess language had been developed 3-4 decades back.

One among them is that children are acquiring more number of skills at an early age. The amount of stimulation along with physiological and psychological maturation is been increased over the year. So the tests may not accurately assess the language abilities in children. Children acquire the skills faster in the present day as compared to the children of the older generation. So the test needs to be revised as well as standardized

2 Aim of the Study

The study is aimed at revising REELS (Receptive Expressive Emergent Language Scale) in Kannada speaking children.

3 Method

720 children within the age group of 0-3 years allover Karnataka were selected for the study. Both male and female children were considered for the study.

Selection criteria
1.For the present study we considered children without any prenatal, perinatal, and postnatal complications.
2.Children not having any behavioral, psychological, physiological, or sensory problems. These children were perceptually evaluated.
3.All children had Kannada (a Dravidian language) as their mother tongue and first language.

Material Used
Developmental milestones specified in Receptive Expressive Emergent Language Scales-REELS [Bzoch and League, 1971] were used as questionnaire for our study. REELS, is a measure of receptive and expressive language skills. The milestones were divided into 3 months intervals ranging from 0-3 months to 33-36 months. The skills mentioned in this test were taken up as questionnaire for both receptive and expressive skills which were numbered and administered.

All children underwent informal screening for Hearing and Vision problems. For "Hearing, informal screening was done using non verbal sounds like 'clap', 'bell', 'knock' Verbal sounds like 'name call' were used. These sounds were presented at 3 feet and at 5 feet distances.
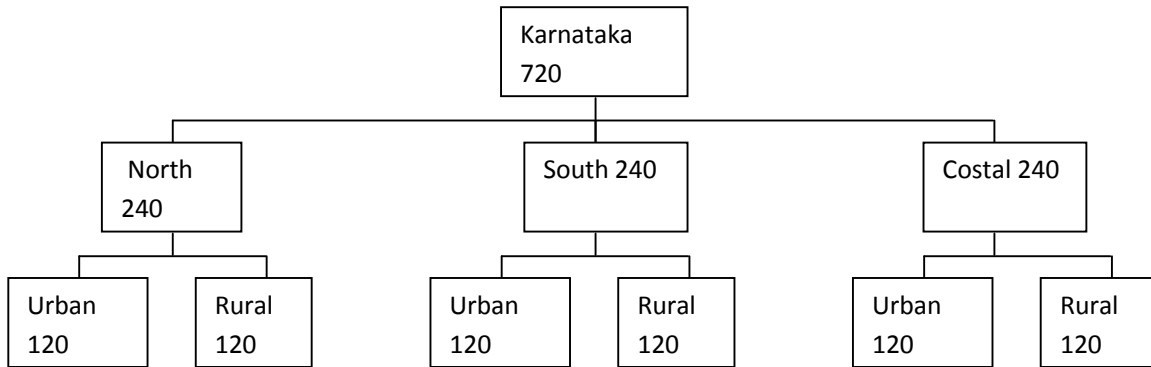
4 Procedure

A total of 720 children in the age range of 0-3 years from all over Karnataka were taken which included both males and females. Data was collected from 3 regions of Karnataka divided as North, South, and Costal regions.

KARNATAKA
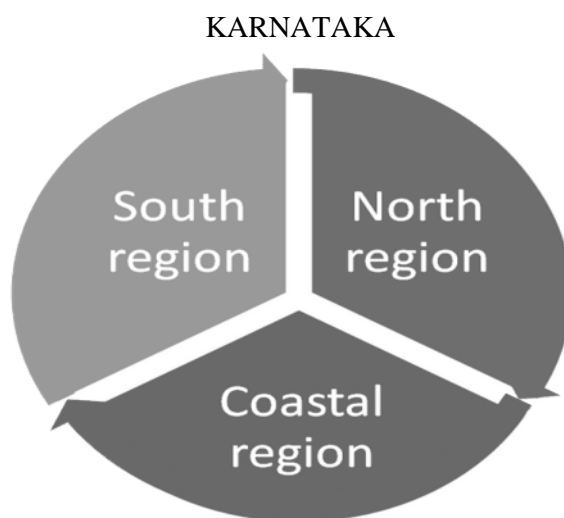
The southern region consists of Mysore, Mandya, Bangalore Urban, Bangalore Rural, Tumkur, Chamarajnagar, Chitradurga, Kolar and Hassan. The coastal region consists of Dakshina Kannada,

Udupi, Uttar Kannada, Chikkamagalur, Karwar and Shimoga. The northern region consists of Bellary, Hubli, Raichur, Gulbarga, Gadag, Haveri, Belgaum, Bidar and Bijapur. In each region children from both rural and urban areas were considered. The samples from North Karnataka include those from North Hubli and Bellary district. From south it includes Mysore, Mandya and Chamarajnagar and from coastal region it includes Dakshina Kannada and Udupi district. Equal number of children from all these regions participated in the study. Children between 0-3 years age range were divided into 12 subgroups with 3 month interval between each consecutive groups.(0-3mnths to 33-36mnths).

```
                            Karnataka
                              720

        North             South 240          Costal 240
         240

   Urban      Rural     Urban    Rural     Urban     Rural
   120        120       120      120       120       120
```

Data was collected at the hospitals, houses; Anganvadi's and play homes, depending on the availability of the children. The data collection began with the Hearing screening followed by administration of the questionnaire to the parents or caretakers.

The questionnaire consisted a total of 86 questions for both receptive and expressive language skills separately (86 x 2 = 172). Questions were asked in the increasing order of difficulty and when the child couldn't do a particular task, the testing was stopped at that point and was noted down in the evaluation form with separate columns for receptive and expressive skills. The speech sample of the children above 2 years of age was recorded. Data entry was in terms of the specific number of questions after which testing was stopped.

5 Result and discussion

The results obtained for all 3 regions were collected and tabulated. The data was arranged according to specific age range starting from 0-3 months till 33-36 months. The mean, standard deviation and t-value was calculated for each of the age groups and the scores obtained in the study were compared with the normative developed in the previous version of REELS (1971). The upper limit at which skills were acquired in each age range was tabulated depending on the number in the questionnaire, where the test was stopped for a particular child. For example, 15-18 months in the previous version of REELS children could satisfy at question number '45' whereas in this study, children were able to satisfy till 54th question.

Table 1: Showing mean, standard deviation and t-value for receptive language skills for both older data and the present one.

Sl no	Age range	Groups	Mean	SD	t- value

1	0-3months	Previous	9	3.6	0.77 ns
		Present	8.25		
2	3-6months	Previous	18	4.29	-0.11 ns
		Present	19.63		
3	6-9months	Previous	27	5.67	0.00 **
		Present	34.17		
4	9-12months	Previous	36	5.1	0.00 **
		Present	41.7		
5	12-15months	Previous	40	7.25	0.02 *
		Present	44.66		
6	15-18months	Previous	45	4.09	0.00 **
		Present	54.0		
7	18-21months	Previous	49	5.18	0.00 **
		Present	57.18		
8	21-24months	Previous	54	2.68	0.00 **
		Present	59.8		
9	24-27months	Previous	57	3.80	0.00 **
		Present	62.31		
10	27-30months	Previous	60	3.90	0.00 **
		Present	65.0		
11	30-33months	Previous	63	3.50	0.00 **
		Present	67.7		
12	33-36months	Previous	66	2.51	0.00 **
		Present	70.0		

(Previous = data of REELS in 1971, Present = data for present population, * = significant, ** = highly significant and ns = not significant)

A Graph 1: Mean difference in Receptive skills between older and present data.
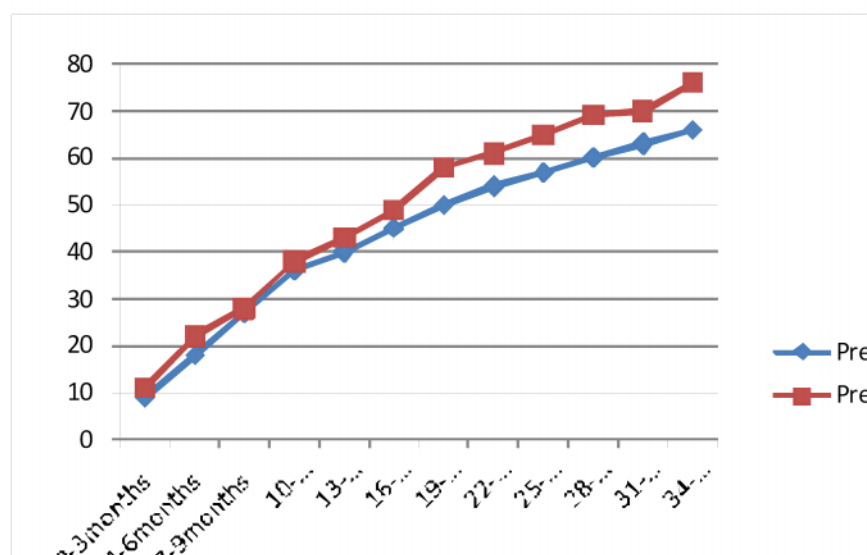
As seen in Table I and Graph I there was a significant difference between the receptive skills achieved from 6 - 36months. No statistically significant difference was found from 0 –6 months. A

significant difference between previous and present data was seen between 1-2 years of age. The difference was comparatively less in the first year as compared to the second year and third year. This shows that children in the present generation acquire skills earlier between 1-2years of age and it increases as the age progresses from 2-3years.

Table 2: Showing mean, Standard deviation and t-value for Expressive language skills for both previous data and the present data.

Sl no	Age range	Groups	Mean	Standard Deviation	T-value
1	0-3months	Previous	9	3.01	0.2 ns
		Present	7.83		
2	3-6months	Previous	18	1.97	0.14 ns
		Present	18.68		
3	6-9months	Previous	27	4.86	0.02 *
		Present	31.23		
4	9-12months	Previous	36	3.67	0.03 *
		Present	38.84		
5	12-15months	Previous	40	3.53	0.05*
		Present	43.0		
6	15-18months	Previous	45	4.25	0.00**
		Present	50.54		
7	18-21months	Previous	49	5.30	0.00**
		Present	55.45		
8	21-24months	Previous	54	3.27	0.00**
		Present	57.9		
9	24-27months	Previous	57	3.86	0.00**
		Present	61.0		
10	27-30months	Previous	60	3.74	0.00**
		Present	64.0		
11	30-33months	Previous	63	2.70	0.00**
		Present	68.0		
12	33-36months	Previous	66	2.11	0.00**
		Present	69.92		

(Previous = data for generation of 1971, Present = data for present population, * = significant, ** = highly significant and ns = not significant).

As seen in table 2, the difference for expressive skills was negligible in the first year of life, but as age progresses, the difference in the skills increased. There is a highly significant difference between the skills acquired at 18-21months to 33-36months. From 6-9 months to 12-15months there is no significant difference between the skills acquisition. This shows that present generation children acquire expressive skills faster in 2-3 years of age as shown in Graph 2.

Graph 2: Mean difference in Expressive skills between older and present data.
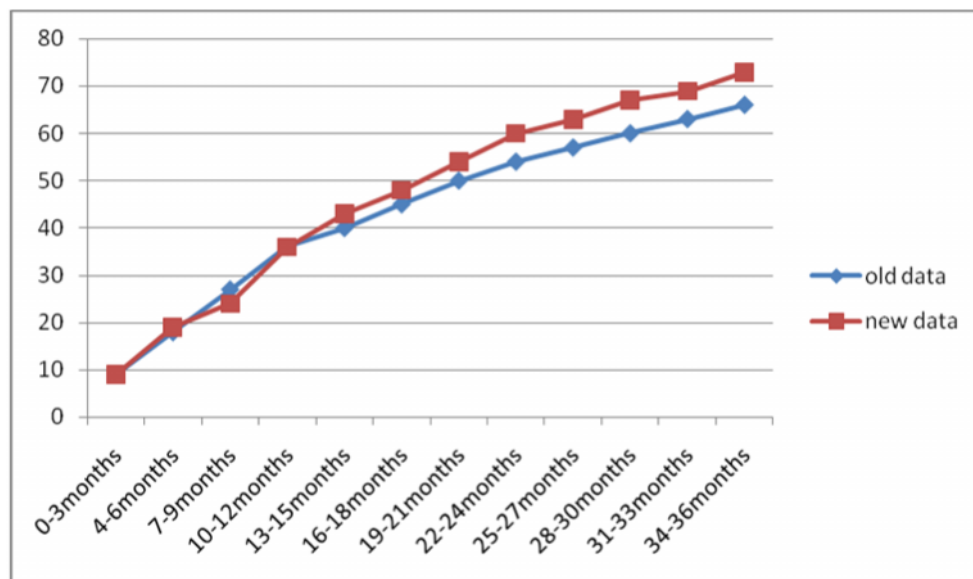
According to this study, skills have improved across the decades, with seemingly high significant difference between the skills acquired by 1-3years of age (Brown and Ellis 1994, Nunes and Bryant 2003). This is because language is another form of learnt behavior, which is acquired as a response to the stimuli in the environment, and then it is learnt. Children's creativity with language and the level of linguistic exposure help them in learning language. Learning is a voluntary response which is strengthened /weakened depending upon positive and negative consequences. These aspects seem to be increasing in the present generations. Parental stimulation and environmental exposure seeming are the important cause for the increased linguistic development for the present generation.

The study revealed that children achieved both expressive and receptive skills in equal amount early in life with reference of REELS I edition developed in 1971 that we revised for the present generation. According to graphs I and II it can be seen that as age progressed there was highly significant difference between older and new data. This difference was minimal during the first year of life but highly significant in the second and third year of life. Hence our aim was to standardize and revise the test. During revision we followed 80 % criteria, that is, if 80% of the children in a particular age group are passing or able to achieve particular milestones, that skill was moved to the lower age group. This was done for both Receptive and Expressive skills. The modified REELS for children who are exposed to Kannada language is as seen in the Appendix. As seen in Appendix the milestones have been shifted to the lower age groups when they were found to be achieved in 80% of the children. The revised version needs to be checked for its validity.

6 Conclusion

This study aimed at standardizing REELS (Receptive Expressive Emergent Language Scale) in the present day Kannada speaking children. The study concluded that there was a significant difference in the receptive and expressive ages. Hence skills that have been achieved early in life have been shifted to the lower age group using 80% criteria. But the scale need to be administered to clinical population and has to be checked for validity.

Acknowledgments

References

Battle. *Communication disorders in multicultural populations*. 3rd ed. Philadelphia: Butterworth, 2002.

Berk. *Child development*. 6th ed. Delhi: Pearson education, 2005.

Bzoch and league. *Receptive expressive emergent language scale*.1971.

Clark. *First language acquisition*. Cambridge: university of Cambridge, 2003.

Catherwood and gillbrand. *Developmental psychology.* Exter: British library, 2004.

Fromkin, Rockman, Hyams. *An introduction to language.* 7th ed. Singpore: Amanda robinson, 2003.

Haynes, shulman. *Communication development: foundations, processes, and clinical applications.* United states of America: Donna balado,1988.

Hulit, Horward. *Born to talk: an introduction to speech and language development.* 2nd ed. United states of America: Stephin D Dragin,1997.

Hurlock. *Developmental psychology. A life span approach*. 5th ed. New York: Elizabeth, 2005.

Luria, Seymour, Trudy. *Language and linguistics in context, reading and application for teachers*. New jersey: Erica kica, 2006.

Mahmud. *Developmental psychology*. New Delhi: Nangia, 2004.

Martin. *speech and language difficulties in the classroom.* 2nd ed . London: caro miller,2003.

Michael, Elizabeth. *Language development, the essential reading*. United states of America: Black well, 2001.

Melaughlin. *Introduction to language development*. 2nd ed. United states of America: Thomson, 2006.

Moore, Peareon. *Competenciesand strategies for apeech language pathology assistance*. Canada: cathy l. erpert, 2003.

Morgon, King, Wiesz and Schopler. *Introduction to psychology*. Richard,

Ramnath and rachana. *Child psychology*. New delhi: 2005

Secord. *Human communication disorders- an introduction*. 4th ed. Place: shames wig.2004.

Shipley , McAfee. *Assessment in speech language pathology A resource manual*. 4th ed. United states of America: Marah Bellegarde, 2009.

Varley. *Introduction to language pathology. 3rd ed. United states of America: crystal, 2000.*

Reflexivity and Causation: A Study of the Vector *'ghe'* (TAKE) in Marathi

Renuka Ozarkar

Department of Linguistics, University of Mumbai, Mumbai

renuka_ozarkar@yahoo.com

Abstract

This paper deals with a compound verb construction in Marathi formed with the vector *ghe* (TAKE), which is ambiguous between a compound verb reading and a causative reading. There is a notion of reflexivity involved in the causative reading of this construction. The paper discusses the distribution and structure of this construction and another common causative construction and attempts to explain their distribution.

1 Introduction

The compound verb construction (V-V complex predicates) is one of the areal features of the South Asian linguistic area. The Indo-Aryan languages in India have a comparably similar set of verbs that occur as vectors or light verbs and form compound verb constructions in these languages. These verbs are typically GIVE, TAKE, COME, GO, PUT, DROP, SIT, etc. These vector verbs show similar semantics in the Indo-Aryan languages in India. They typically contribute some aspectual information such as telicity / perfectivity or inception (Butt 1995, Ramchand 2008). They also add some semantic nuances to the compound verb construction such as benefaction, inadvertency, direction of the action, etc. These nuances are also comparably similar in these languages (Kachru 1993). The vector TAKE in these languages generally expresses a reflexive sense (Masica 1976, Wali 2006). The term reflexivity has not been used in the sense of the syntactic reflexivity, where an argument of the verb gets absorbed or the arguments are co-indexed to each other. In this paper, this term refers to the situation in which the direction of the action denoted by the main verb is towards the agent, or the result of the action is transferable to the agent.

The vector *ghe* (TAKE) in Marathi is also used to form causatives from some verbs. In most cases, the verb + *ghe* construction is ambiguous between the compound verb reading (only reflexive and completive sense) and the causative reading. However, even in its causativising usage, the vector retains its reflexive semantics. That is, the causer is somehow involved in the main action or is the beneficiary / recipient of the result of the caused action. For example,

1. mii   rameS-kaDuun   aambe   paaD-uun   ghet-l-e
     I   Ramesh-from   mangoes   drop.trans-CP   TAKE-perf-3p.pl.masc.

I got the mangoes dropped by Ramesh. (i.e. I caused Ramesh to take the mangoes down from the tree).

2. panCam-ne   he   gaaNe   giitaa-kaDuun   gaaw-uun   ghet-l-e
     Pancham-ERG   this   song   Gita-from   sing-CP   TAKE-perf-3p.sing.neut

Pancham got this song sung by Gita.

In (1), the causer 'I' is the recipient or the beneficiary. In (2), the causer is closely involved in the action of singing: he instructed or closely monitored the act of singing. He is also the beneficiary of the act of singing, in some sense. In other words, the reflexivity and the causative sense of the vector *ghe* appear to be closely linked.

Use of the vector *ghe* to form causatives has gone largely unnoticed in the literature on Marathi (except Shibatani & Pardeshi, 2003). In this paper, an attempt is being made to discuss the structure and the distribution of the vector *ghe* in its causativising usage and to analyze the link between its

reflexive and causative semantics and to capture. For the purpose of the analysis, Ramchand's decompositional framework of the first phase syntax is used, which seems to capture the ambiguity between the compound verb reading and the causative reading of *ghe*.

Section one of the paper consists of a brief overview of the causativization strategies in Marathi. It also deals with the distribution of these strategies, which include causativization by *ghe*. Section two deals with the analysis of the causative constructions in general and of the *ghe* causative in particular. I will also attempt to explain the distribution of the *ghe* causative in this section. The next section addresses the issue of the reflexive sense of the *ghe* causative. However, my analysis is still under construction and so I will only mention a tentative explanation of the reflexivity in the *ghe* causative.

2 An overview of the causatives in Marathi

Traditional grammars of Marathi (Damale, 1969 [1911]) mention more than one strategy in Marathi to causativize the verbs. Shibatani & Pardeshi (2003) discuss the following as some of the 'synthetic' causatives in Marathi:

   i. Internal consonant change

   phaaT-Ne (to get torn) *: phaaD-Ne* (to tear)

   ii. Internal vowel change

   mar-Ne (to die) *: maar-Ne* (to kill)

   iii. Internal consonant and vowel change

   phuT-Ne (to burst) *: phoD-Ne* (to burst trans.)

   iv. Suffixation, i.e. adding –aw suffix to the root:

   has-Ne (to laugh) *: has-aw-Ne* (to make laugh)

Apart from these, there are pairs such that one of the pair is semantically causative of the other, but there is no similarity in their roots, so that both of them are treated as distinct verbs (Damale 1969[1911]). For example, *dzaa-Ne* (to go): *paaTh-aw-Ne (*to send), *je-Ne* (to come): *aaNa-Ne* (to bring). Shibatani & Pardeshi call these as suppletives. There is also a set of labile verbs, i.e. the verbs, which can be used both as intransitive and transitive. For example, *ughaD-Ne, ukaL-Ne, ukaD-Ne, moD-Ne.*[1]

There are 'analytic' causative forms as well in Marathi. Shibatani & Pardeshi mention the following:

   i. *laaw* (lit. attach, apply):

   3. aaii-ne   raam-laa   gaaNe gaay-laa laaw-l-e

   mother-ERG Ram-DAT Ram-DAT song sing-DAT MAKE-perf-3p.sing.neut.

   Mother made Ram sing a song.

   ii. *bhaag paaD-Ne*[2]:

   4. aaii-ne   raam-laa   gaaNe gaay-laa bhaag paaD-l-e

   mother-ERG Ram-DAT song sing-DAT MAKE-perf-3p.sing.neut.

   Mother made Ram sing a song.

   iii. *de-Ne* (give) :

   5. aaii-ne raam-laa gaaNe gaa-uu di-l-e

   Mother-ERG Ram-DAT song sing-P GIVE-perf-3p.sing.neut.

---

[1]   Damale also mentions some transitive verbs, which although look like –aw causative forms, are not causative forms of any existing verbs in Marathi. For example, *bolaaw-Ne (to call, to invite).*Although there is a verb *bol-Ne* (to speak), which seems like the non-causative form of *bolaaw-Ne,* synchronically ate least it cannot be regarded as such.

[2] Following Shibatani & Pardeshi, I gloss this verb as 'MAKE' (block letters mine). They also give the literal translation as 'make fall in one's destiny'.

Mother let Ram sing a song.
  iv. *ghe-Ne* (take):
    6. aaii-ne raam-kaDuun gaaNe gaa-uun ghet-l-e
      mother-ERG Ram-from song sing-CP TAKE-perf-3p.sing.neut.
      Mother got a song sung by Ram.

The first two of the analytic causatives have a coercive reading. The *de* causative is called a permissive construction. For the purpose of this paper, I will keep (i-iii) of the analytic type out of the discussion and focus only on the *ghe* causative. I will also club (i-iii) of the synthetic type together and refer to them as forms obtained by root-alternation.

Not all strategies in Marathi apply to all types of verbs. So let us take a glance at the distribution of the causative constructions in Marathi.

2.1 The causative constructions and the verb classes

In order to understand the distribution of the causativizing mechanisms in Marathi, it is necessary to consider the classes of verbs. Broadly, let us consider the transitive verb class and two intransitive verb classes, namely unaccusatives and unergatives.

It is found that only verbs from the unaccusative class have corresponding causative forms by root-alternation. It is a highly restricted, closed set of unaccusative verbs that have such transitive counterparts. Following Ramchand, I treat them here as base transitives. These transitive verbs indeed show similar behaviour as regular transitives.

Among the synthetic group, the –aw causatives are most productive. This suffix causativises most of the unergative verbs as well as many of the unaccusative verbs. However, it does not causativise transitive verbs. Very rarely, it does causativise *kar-Ne* (to do) as *kar-aw-Ne* and *lih-Ne* (to write) as *lih-aw-Ne.* But these too are not very common instances.[3]

Transitive verbs are generally causativized by using the *laaw* verb. This causativising mechanism is highly productive in Marathi and it takes place in case of all types of verbs except the unaccusative verbs.

The *ghe* causative takes place only in case of the transitive verbs. The vector *ghe* also causativises the transitive counterparts of the unaccusative verbs. It, however, causativises only some of the –aw causatives of the unaccusative verbs and does not causativize most of the –aw causatives of the unergative verbs. There are some exceptions to this general pattern, but our analysis will be able to explain them in the end.

The distribution of the synthetic causatives under consideration here can be summarized as in the table (1). We also take a look at the distribution of the vector *ghe* (non-causative reading) in order to compare it with the causative use of *ghe* in table (2).

	Internal change (base transitives)	-aw suffixation
Unaccusative verbs		
Unergative	X	

---

[3]   The suffix –aw must have been very productive earlier in Marathi. In Damale's grammar, -aw is said to regularly apply to the transitive verbs as well. He goes further to say that this suffix also forms 'secondary causatives' from the transitive counterparts of the unaccusative verbs. For example, paD-Ne : paaD-Ne : paaD-aw-Ne. Some speakers of Marathi today may use this particular form of the verb. However, the productivity of this suffix in general seems to have reduced since Damale's time. Today, this suffix is largely restricted to intransitive verbs.

verbs		
Transitive verbs	X	X

Table (1): distribution of synthetic causative

	ghe vector	*ghe* causative
Unaccusative verbs	X	X
Base transitives of unaccusatives		
-aw causatives of unaccusatives		
Unergative verbs		X
-aw causatives of unergatives		X
Transitive verbs		

Table (2): distribution of *ghe*

As said earlier, I treat the transitivised forms of the unaccusative verbs obtained by root-alternation as base transitives. So, now on our focus is mainly on the distribution and structure of the –aw causatives and the *ghe* causatives. Apart from the fact that certain –aw suffixed verbs can further be causativised by *ghe* vector, there is also a semantic difference between the two forms: the –aw suffixed causative forms seem to give the sense of direct causation, while the *ghe* causatives express indirect causation.

2.2 Direct and indirect causation

The –aw causative gives a direct causative sense as against the *ghe* causative. That is, the agent of the non-causative verb, when becomes a causee in –aw causative, it behaves like a patient. The DP denoting this causee also takes the dative/accusative case marker. For example,

7. a) baaL dzhop-l-e
     child sleep-perf-3p.sing.neut.
     The child slept.
  b) aaii-ne   baaLaa-laa dzhop-aw-l-e
     mother-ERG child-ACC sleep-caus.-perf-3p.sig.neut.
     The mother made the child sleep.

The causee in (7.b), i.e. the child, undergoes the activity of sleeping. There is a single agent/causer i.e. the mother in this case and no intermediate agent is possible. Consider (7.c)

  c) */ ? aaii-ne   raam-kaDun baaLaa-laa dzhop-aw-l-e
     mother-ERG Ram-from   child-ACC sleep-caus.-perf-3p.sing.neut.
     The mother made the child sleep through Ram.

Some of the speakers may find this acceptable if instead of the postposition –kaDun, other postpositions like –karavii (through), -maarphat (through) are used.

In case of the *ghe* causatives, however, there is a sense of indirect causation. The causer may not directly bring about the action, although he/she is very much involved in the activity. There is a possibility of a more directly involved intermediate agent, which is expressed by the adjunct DP taking the postposition -kaDun. This intermediate agent is the agent in the non-causative counterpart. In other words, the non-causative agent does not act like a patient in its causative form, but is instead an intermediate agent. It retains its initiational properties. Consider (8).

8. a) baaii-ne   kapaDe dhut-l-e
     maid-ERG clothes wash-perf-3p.pl.musc.
     The maid washed the clothes.
  b) siimaa-ne   (baaii-kaDun) kapaDe dhuw-uun ghet-l-e
     Seema-ERG maid-from   clothes   wash-CP   TAKE-perf-3p.pl.masc.
     Seema got the clothes washed by the maid.

The agent of the non-causative form, i.e. the maid, does not act like a patient and does not take the dative/ accusative case marker.

Ramchand (2008) and Shibatani & Pardeshi (2003) independently explain the direct and indirect causation in terms of the event structure. Ramchand claims that if the process and result are temporally adjoining, i.e. the process leads to the result without a time gap between them, then the result is 'direct'. If not, then the result is 'indirect'. In her framework, Ramchand assumes the relation between the initiation subevent and process subevent as of causation, as is between process and result subevents. If the process and result subevents are causally related, as they always are, but temporally distinct or independent, then we get the indirect causation sense. There is a potential for an intermediary (Ramchand, 2008: 170). The temporal independence between the two subevents is obtained if those subevents are identified by distinct lexical items. She puts this claim in a hypothesis as follows (Ramchand, 2008: 170):

The Temporal Dependence Hypothesis

For a result subevent to be temporally dependent on a process, the same root must identify the two subevents.

So we expect that in case of the –aw causatives in Marathi the proc and res are identified by the same item, while in case of the *ghe* causative, they are identified by different items. Let us see how the structures look like.

## 3 Structures of the –aw and *ghe* causatives in Marathi

The unaccusative verbs lack init feature. They are [proc, res] type. And their only argument is UNDERGOER-RESULTEE. In case of the transitive counterparts of these verbs, the init feature is also lexicalized and such verbs have two arguments: one is INITIATOR, the other is UNDERGOER-RESULTEE. If the –aw is added to the unaccusative verbs, then –aw lexicalizes the init feature and contribute a corresponding INITIATOR argument. No subevent is underassociated.

The unergative verbs are typically [init, proc] type. Their only argument is INITIATOR-UNDERGOER. When such a verb is causativised by –aw suffixation, its init becomes underassociated and is in turn, lexicalized by –aw. It adds the INITIATOR argument. The original argument of the unergative verb remains in the UNDERGOER position. That gives the patient-like properties to that argument (the causee).

In both unergative and unaccusative verbs, what would be a causee in their –aw causativised forms, is also an UNDERGOER. In case of the transitive verbs, there are two arguments, one that is an INITIATOR and another that is the UNDERGOER. **The –aw causative would require that the causee would be the UNDERGOER, which is not possible in case of the transitive verbs. This can explain why, -aw causative is not possible for the transitive verbs.** The –aw causative gives a direct causation sense, for which it is required that the causee acts like a patient or undergoer. The causee is directly, at times even physically affected by the act of causation. **In order to satisfy this requirement of the –aw causatives, it is necessary that the INITIATOR, if present, must also be the UNDERGOER. This is possible only in case of unaccusative, which have only UNDERGOER argument and unergative verbs, which have the same argument being both the INITIATOR and the UNDERGOER, but not in case of transitive verbs.**

162

However, the transitive verbs can be causativised by using the vector *ghe.* Referring back to the table (2), it is clear that the *ghe* causative is possible only in case of those verbs, which can form compound verbs with the vector *ghe.* It has also been noted earlier that this kind of causative retains the reflexive sense that the vector *ghe* adds to the compound verb construction. I take this as a lead to assume that the causative *ghe* and vector *ghe* as in a CV have the same structure and that they hold similar selectional properties. In fact, in case of the transitive verbs, the vector *ghe* is ambiguous between the vector reading and the causative reading. Following Ramchand's decompositional analysis of the CV constructions, the structure of (9.a), which is an example of a CV, will be as in (9.b).

9. a) siima-ne   sagaLe kapaDe dhuw-uun ghet-l-e
    Seema-ERG all   clothes wash-CP   TAKE-perf-3p.pl.musc.
    Seema washed up all the clothes.

9. b)
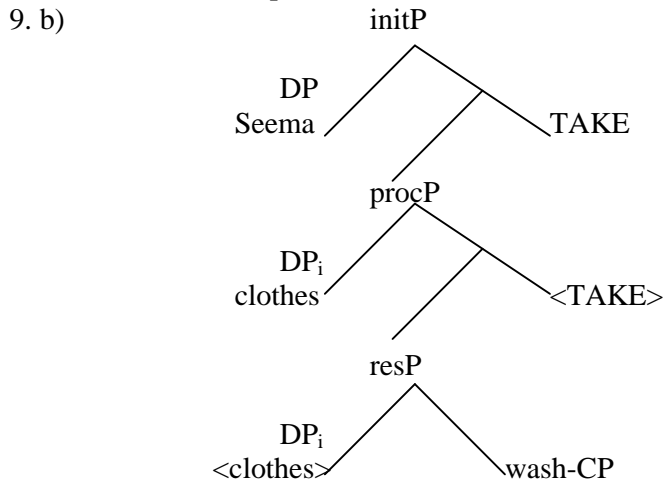
Here, the vector lexicalizes both the init and proc features, while the participial form of the main verb lexicalizes the res feature. If we assume the same structure for the *ghe* causative, it generates the indirect causation reading, as the proc and the res features are lexicalized by different lexical items. This is the desired effect.

Lexicalization of the proc and res features by different lexical items represents the temporal independence between the subevents as expected and thus generates the possibility of the intermediate agent. Interestingly, although the intermediate agent DP is an adjunct, it is required in order to disambiguate the causative reading. Without this DP, the most prominent interpretation of the v-uun+*ghe* construction will be that of the compound verb. So, the example (9.a) is a CV construction and Seema is interpreted as the direct agent of the act of washing the clothes. However, there is a possibility of adding the intermediate agent DP, say *baaii-kaDun,* as in (8.b). In this case, the construction is read as a causative construction and Seema is no longer the direct agent. Interestingly, the difference in the two interpretations is reflected in the entailments of each one. While (9.a) entails that Seema washed the clothes, (8.b) does not entail so. It indeed entails that the maid washed the clothes. Given these facts, the status of the intermediate agent DP needs a closer look.

Looking back to the table about the distribution of the causative *ghe,* we find that this verb also causativizes the transitivised unaccusative verbs. Consider:

10) raam-ne   (madzuraan-kaDuun) dzhaaDe toD-uun   ghet-l-ii
    Ram-ERG labourers-from   trees   break.trans-CP TAKE-perf-3p.pl.neut
    Ram got the trees cut by the labourers.

This is not surprising, as we are treating these verbs as base transitives. However, *ghe* can further causativise even the –aw suffixed causatives of the unaccusaive verbs. Interestingly, it does not causative the –aw suffixed causatives of the unergative verbs. But, there are some exceptions to this general observation. Consider (11), in which the vector *ghe* causativises the –aw causative of the unergative verb *bas-Ne.*

    11) aaii-ne   (raam-kaDun) bhintii-war aarsaa bas-aw-uun
       Mother-ERG Ram-from   wall-on   mirror sit-caus-CP
       ghet-l-aa
       TAKE-perf-3p.sing.musc.
     The mother made Ram fix a mirror on the wall.

Here we have to account for two facts: (1) the vector *ghe* causativises already causativised –aw suffixed forms, and (2) there are restrictions on this, i.e. the vector does not causativise all such forms.

The direct and indirect causative readings may prompt us to think that the vector *ghe* is obtained by the recursion of the verb-layer in syntax. However, we showed that *ghe* causative has the same structure as the compound verb. And the compound verbs in Marathi are monoclausal, i.e. they do not involve any recursion of the verb-layer. Also, we noted that the indirect sense of the *ghe* causative comes not from any recursion but from the fact that the proc and res subevents are temporally independent. So, we will handle the fact (1), through the underassociation of features. I suspect that in her framework, Ramchand would also handle the fact that the Hindi –vaa causative can further take a vector to form a compound verb in similar way.

4 Explanation of the distribution of the *ghe* causative

As we know, -aw identifies the init feature and contributes the causer-INITIATOR argument to the construction. When such a construction further takes the vector *ghe,* the init and the proc features of the causativised verb are underassociated and are subsequently identified by the vector *ghe.*[4]

In case of the unaccusative intransitive verbs, -aw suffix identifies the init feature and contributes the INITIATOR argument. We noted that the causativising vector *ghe* could occur with those verbs, which form a CV with the vector *ghe.* Now to be able to form a CV with this vector, the verb must have init feature in the first place, which will be underassociated. The unaccusaive verbs do not have init feature, and thus do not take the vector *ghe* at all. The –aw causatives of these verbs, however, identify the init feature thereby, permitting the *ghe* vector and its causative reading.

In case of the unergative verbs, however, this feature is already present and thus they take the vector *ghe.* But, they do not get the causative reading. As noted before, for the causative reading of the vector, it is essential that the UNDERGOER argument should not have any initiational properties, and that the UNDERGOER and INITIATOR arguments are to be identified by separate entities. It does not happen with the unergative verbs. Although, -aw suffix now identifies two distinct UNDERGOER and INITIATOR arguments, the UNDERGOER still has initiational properties such as control over the action it is undergoing. Thus, the condition now being satisfies, the both the unergative verbs and their –aw causatives do not permit the *ghe* causatives.

However, in case of the exceptions, we see that the UNDERGOER arguments of the unergative verbs are such that they do not have any control over the action or they do not have volition. Such

---

[4]   At this stage, it is not clear whether the –aw suffixed form is to be taken as a base form or to be built in syntax. I feel, contrary to Shibatani & Pardeshi that –aw causative, although restricted, is productive and regular enough to be handled in syntax. At least, I do not find enough motivation for it to be considered as a base form. But I leave this issue to further probe.

an UNDERGOER does not have initiational properties. We can test this by using adverbs of deliberation with the unergative verbs.

12.a) siimaa muddaam   caTai-war bas-l-ii
  Seema deliberately mat-on   sit-perf-3p.sing.fem.
  Seema sat on the mat deliberately.

  b) *aarsaa muddaam   bhintii-war bas-l-aa
  mirror deliberately wall-on   sit-perf-3p.sing.masc.
 The mirror sat/ fitted on the wall deliberately.

In case of (12.a), the *ghe* causative is not possible, as shown in (12.c) below. However, in case of (12.b), it is possible to form the *ghe* causative (12.d).

  c) *aaii-ne   raam-kaDun siimaa-laa   caTai-war bas-aw-uun
   mother-ERG Ram-from   Seema-ACC mat-on   sit-caus.CP
   ghet-l-a
   TAKE-perf-3p.sing.neut.
   The mother made Seema sit on the mat through Ram.

  d) aaii-ne   (raam-kaDun) bhintii-war aarsaa bas-aw-uun
   Mother-ERG Ram-from   wall-on   mirror sit-caus-CP
   ghet-l-aa
   TAKE-perf-3p.sing.masc.
   The mother made Ram fix a mirror on the wall.

In case of the transitive verbs, the init feature is identified by the verb. The INITIATOR and UNDERGOER arguments are distinct entities. Also as in case of (9), the UNDERGOER has no initiational properties. Thus all conditions for the occurrence of the vector *ghe* and its causative reading are satisfied.[5]

To summarise, I put the conditions required for the occurrence of the causativising vector *ghe* as follows:

The non-causative verb must have identified the init feature, which will be subsequently underassociated.

A) The INITIATOR and UNDERGOER arguments must be distinct entities.

B) The UNDERGOER argument must not have initiational properties such as volition or control over the activity that it is undergoing.

The vector *ghe* in its causativising sense occurs only with those verbs, which satisfy all these conditions. The overall picture can be represented as in the table (3).

No.	Verbs	Condition 1	2A	2B	*ghe* vector	*ghe* causative
1.	Unaccusative	X	X		X	X
2.	Transitive forms of unaccusative verbs (internal change)					

---

[5]   This shows that –aw suffix is not required in order to causativise a transitive verb by *ghe*. It is already capable of taking this causativising vector. Earlier the -aw suffix used to be highly productive and used to causativise even the transitive verbs. Sometimes, the combinations such as *kar-aw-uun ghe-Ne* do occur in Marathi, even today. But, this is to intensify the causative sense. Also such instances are rare and most of the transitive verbs do not have –aw suffixed causative forms in the language anymore. Our analysis captures the fact that if the vector *ghe* is added to the –aw suffixed transitive verb, then this suffix becomes syntactically completely redundant. The INITIATOR argument originally contributed by the –aw suffix, now gets reinterpreted as the contribution of the vector itself. Thus the vector becomes the causativising element. The –aw suffix, due to other reason, falls out of usage in the domain where it is functionally redundant. This is a tentative historical account. However, at present I have no evidence to attest it. Nonetheless, it is an interesting speculation.

3.	-aw causative forms of unaccusative verbs					
4.	Unergative verbs		X	X		X
5.	-aw causative forms of unergative verbs			X		X
6.	-aw causative forms of unergatives (eg. aarsaa basawNe)					
7.	Transitive verbs (eg. kaam karNe, gaaNe gaaNe)					
8.	ukaDNe, ughaDne, moDNe (intransitive)	X	X		X	X
9.	ukaDNe, ughaDne, moDNe (intransitive)				X*	

* in this case, the vector *ghe* does not get the CV reading, because the V+*ghe* form will be identical for both transitive and intransitive verbs of this group. And such CV construction is not possible in case of intransitive form. Thus the only reading we get is that of causative.

Table (3): Verb classes and conditions for causatives

However, consider (14), in which all these conditions are satisfied, but the vector's causative sense is blocked.

14) *aaii-ne   baabaan-kaDun raam-laa   Cop-uun ghet-l-a
     Mother-ERG father-from   Ram-ACC beat-CP   TAKE-perf-3p.sing.neut.
     The mother got Ram beaten up by the father.

And here, the notion of reflexivity as we used it earlier, or the notion of agent/self-benefaction comes into picture.

5 Reflexivity

For the felicity of the *ghe* causatives it is necessary that the object/ UNDERGOER to be transferable to the agent. At least the effect of the action should be transferable. This is what we have called as reflexivity sense of the vector *ghe*. How to get this condition satisfied is still a puzzle for me at the moment. Both the conditions mentioned above together may help to get the desired result. First, the UNDERGOER is not the same as the INITIATOR and has no initiational properties. It has no control over the action. Secondly, given the structure of the *ghe* causative, it can be seen that the UNDERGOER is the subject of the head lexicalized by the vector *ghe*. The same argument is also the RESULTEE. Such a construction may give the effect of 'argument-sharing'. Note that in a CV construction too, the vectors lexicalize the head whose subject is the

UNDERGOER/object as well as the RESULTEE, wherein the res head is lexicalized by the main verb. This effect of argument-sharing may help to obtain the reflexive sense.

However, I suspect that this is not sufficient and the reflexivity sense of the *ghe* causative may be handled in the matching of conceptual-encyclopedic content of the verb and the vector *ghe*. Consider examples like (15) and (16). The reflexive sense is more metaphorical in these cases and cannot be easily explained by simply the notion of transferability or the benefaction of the agent.

15) aapaN tiCyaa-kaDuun yaaCii   tayaarii   karuun ghe-uu
    we   she-from   his(prox) preparation do-CP TAKE-fut.-1p.pl
    We will get this one prepared by her.

16) aaii-ne   madZhyaa-kaDuun kavitaa paaTh   mhaN-uun
    mother-ERG   me-from   poems   by-heart say-CP
    ghet-l-yaa
    TAKE-perf-3p.fem.pl.
    Mother got the poems sung by heart by me.

In either case, the causer is not the beneficiary or the recipient. It is more metaphorical than the earlier examples.

Nonetheless it will be interesting to look at what constitutes the two contents and what makes their matching possible or bars their matching. My analysis is still under construction and so for the time being I will leave it at this tentative proposal.

6 Conclusion

In this paper, I attempted to probe into the relationship of the vector *ghe* (TAKE), its causativising function and the sense of reflexivity that it has. In order to understand this relationship, I discussed the structure and distribution of the vector *ghe* and its causativising sense.

Given the ambiguity between the compound verb reading and the causative reading of the vector *ghe*, I take its structure as same as that of the compound verb formed with this vector. In such a case, the vector identifies the init and the proc features. The res feature is identified by the verb root. This structure gives the indirect causative sense expected in the *ghe* causative construction.

The occurrence of the vector *ghe* in its causativising sense is determined by the following conditions.

The non-causative verb must have identified the init feature, which will be subsequently underassociated.

A) The INITIATOR and UNDERGOER arguments must be distinct entities.

B) The UNDERGOER argument must not have initiational properties such as volition or control over the activity that it is undergoing.

C) The UNDERGOER entity should be transferable to the agent-causer. At least the result should be transferable. This condition is required to get the reflexive sense of the *ghe* causative.

The first two conditions are handled in syntax. The third one, it seems, is to be handled by the unification of the encyclopedic-conceptual content of the main verb and the vector.

This analysis, based on Ramchand's decompositional framework of the first phase syntax, captures the regularity in the distribution of both the –aw causative and the *ghe* causatives in Marathi. It also shades light on the interaction of these two causative constructions.

References

Damale, Moro Keshav. 1969. *Shastriiya Marathi VyaakaraN*. Ed. Krishna Arjunwadkar. Pune: Deshmukh & Company.

Ramchand, Gillian. 2008. *Verb Meaning and Lexicon: The First Phase Syntax.* Cambridge: Cambridge University Press.

Shibatani & Prashant Pardeshi. 2003. 'The Causative Continuum'

http://www.lit.kobe-u.ac.jp/linguistics/KPL/3_2001/KPL_2001_shibatani-pardeshi.pdf

Wali, Kashi. 2006. *Marathi.* Indian Institute of Languages.

Cross language variations in linguistic deficits in Dementia of Alzheimer's type (DAT) individuals

Sunil Kumar Ravi
rsunilkumar86@gmail.com

Shyamala. K. Chengappa
shyamalakc@yahoo.com

Department of Speech – Language Pathology,
All India Institute of Speech and Hearing,
Mysore, Karnataka, INDIA – 570006

Abstract

Dementia is a common clinical syndrome characterized by a decline in the cognitive function and memory from the previously attained intellectual levels, which is sustained over a period of months or sometimes even years. The aim of the present study is to study the cross language variations in language deficits in DAT individuals in English and Kannada. Around 20 individuals with DAT in the age range of 45 to 65 years were taken as subjects for this study. The Performance of the individuals with DAT in Kannada and English were compared to study the cross language variations in these subjects. And also the results of the individuals with DAT were compared with controls. Our Results indicated significant differences in performance of individuals with DAT between two languages. It was found that the individuals with DAT showed better performance in the most used language than the other language.

1 Introduction

Dementia is a common clinical syndrome characterized by a decline in the cognitive function and memory from the previously attained intellectual levels, which is sustained over a period of months or sometimes even years. The deterioration is of such severity that it impairs the affected individual's ability to work and to perform activities of day to day life, including communication.

The language disturbances in dementia have long been reported. Interest in the dementias has increased in the past decade resulting in a more systematic description of the effects of dementia on communication. This produced not only a more fundamental understanding of the disorder, but improved avenues of management. Since dementing illnesses are associated with the elderly, the expectation and unavoidable conclusion is that the prevalence of dementia will increase.

Dementia can be caused by a variety of conditions: diseases, infections, or infarcts. The most commonly occurring cause is Alzheimer's disease accounting for 50 to 60% of all the patients with dementia. Vascular dementias (dementias caused by multiple infarcts) are seen in 20% of the dementia patients. Alzheimer's dementia and vascular dementia co-occur in approximately 15% of this sample, and other conditions such as Pick's disease, Parkinson's disease, progressive supranuclear palsy, and creutzfeldt-Jacob disease, account for the reminder of the irreversible dementias.

2 Dementia of Alzheimer's Type (DAT)

Dementia of Alzheimer's Type (DAT) refers to a type of dementia which is associated or caused due to Alzheimer's disease. Alzheimer's disease (AD) is the most common cause of dementia, affecting as many as 4 million Americans. AD is a degenerative disease that attacks the brain, begins gradually, and progresses at a variable rate. AD results in impaired memory, thinking, and behavior and can last from 3 to 20 years from the time of the onset of symptoms.

Warning signs of AD are memory loss that affects job/home skills, difficulty performing familiar tasks, problems finding the right words, disorientation as to time and place, poor or decreased judgment, difficulty with learning and abstract thinking, placing things in inappropriate places, changes in mood and personality, and marked loss of initiative. In the last stage of AD, patients are unable to take care of themselves. Recent research has shown links between particular genes and Alzheimer's disease, but in about 90% of AD cases, there is no clear genetic link. With the help of standardized diagnostic criteria, physicians can now diagnose AD with an accuracy of 85-90% once the symptoms occur. However, a definitive diagnosis of Alzheimer's disease is possible only through the examination of brain tissue at autopsy.

AD is a debilitating progressive disease which gradually affects all aspects of cognition and behavior, including aspects of semantic memory and semantic knowledge (Hodges and Patterson, 1995). Most patients with confirmed Alzheimer's disease appear to have fluent speech with poor semantic content and comprehension (Cummings *et al.*, 1985). Selective loss of the appropriate use and recall of word meanings and object naming and recognition occurs particularly in discourse (Bayles *et al.*, 1993) while phonology (sound structure), syntax (sentence order and structure) and grammar are usually well preserved until later in the disease process. The brain substrate of the seemingly selective and specific loss of types of language functions in Alzheimer's disease is intriguing.

Dementia of Alzheimer's Type in Bilinguals

Studies have suggested that bilingual speakers with AD, even in the early stages of deterioration, can have problems selecting the appropriate language, and maintaining conversations in that language once appropriately chosen (De Santi *et al*. 1990 & Luderus 1995).

Following this distinction of language choice versus language separation problem, most of the few studies on bilingual AD have investigated discourse management along these lines (De Santi *et al*. 1990, Luderus 1995). Researchers set out to answer questions such as whether a given speaker has either a choice or a separation problem; whether every bilingual speaker with AD evidences these difficulties; how the incidence and prevalence of a choice or separation problem relates to severity of dementia; whether the direction of mis-selection or intrusion of one language into another is bidirectional and unpredictable, or whether, for instance, is it always one language (e.g. the most proficient/ the first learned/ the most recently used) that takes precedence over the other; and so forth? Conclusions indicate that not every bilingual individual with AD demonstrates inappropriate language use (e.g. Luderus 1995). There is a large variability in the extent of appropriate/ inappropriate language use, with some individuals showing more language mixing than others.

Therefore, there is a great need to study the cross language variations in bilingual dementias to explore the nature of language deficits in each language which will in turn help in both assessment and management of these individuals.

3 Aim of the Study

The aim of the present study was to explore the nature of the linguistic deficits in both the languages of bilingual DAT individuals and also to study the differences in performance between bilingual DAT individuals and normal population.

4 Method

Subjects

20 individuals who were diagnosed as Dementia of Alzheimer's Type by a neurologist or psychiatrist in the age range of 45 to 65 years were taken as the subjects for this study. The following inclusion criteria was used to select the subjects for this study: (1) subjects who are diagnosed as having DAT by a neurologist or psychiatrist, (2) subjects who have Kannada as their native language (L1) and English as their second language (L2), (3) a relatively stable clinical and metabolic state, and (4) a fair knowledge of reading and writing in Kannada and English languages. Twenty age and gender matched normal subjects without any past neurological or psychiatric disorders were taken as the control group.

Test material

All the subjects were assessed on Dementia Assessment Battery (DAB) in English and Kannada which was developed for this study. DAB consists of four domains, memory, linguistic expression, linguistic comprehension, visuospatial skills with each domain consisting of several subtests. Table -1 gives an overview of the subtests of DAB.

	Memory	Linguistic Expression	Linguistic Comprehension	Visuospatial skills
1	Episodic Memory	Object Naming	Comparative Questions	Generative Drawing
2	Working Memory	Generative Naming	Following Commands	Figure Copying
3	Semantic Memory	Confrontation Naming	Reading Comprehension of Sentences	
4	Delayed Story Telling (Wh – questions)	Responsive Speech	Reading Commands	
5		Spontaneous Speech		
6		Repetition		

Table – 1: List of domains and subtests of each domain of DAB.

Data Collection

Data collection was done at the Department of Speech – Language Pathology at the All India Institute of Speech and Hearing, Mysore, India with the subjects hailing from Mysore and Bangalore cities.

5 Results and Discussion

As the aim of the present study was to explore the nature of language deficits in individuals with DAT and also to explore the cross language variations in bilingual DAT individuals, the results of various tasks like memory, language expression and language comprehension were compared between DAT individuals in Kannada and English languages and the results were also compared with that of results of normal adults. Majorly, four comparisons were made in this study to find the cross language variations in DAT individuals.

    i.   **Comparison of performance of normal adults in Kannada and English Languages**

Comparison of performances was done using independent samples t – test to find out the significant difference between two groups across the parameters. Statistical analysis showed no significant difference between the two groups on memory (p>0.005, f = 0.10) and language expression (p>0.005, f = 0.08) tasks and showed significant difference on language comprehension tasks (p<0.005, f = 0.01). This suggests that normal adults performed very similarly in Kannada and English on memory and language expression. A slight difference was observed in the performance of normal adults on language comprehension tasks in English and Kannada and this can be attributed to the amount of language exposure and the age of acquisition of second language of the normal group. The mean and S.D of the two groups are given in Table – 2. The mean and S.D of the normal individuals is greater in Kannada language than in the English language.

Tasks	Group	N	Mean	Std. Deviation
Memory	Normal Kan	20	86.5500	2.4810
	Normal eng	20	81.3500	4.9553
Expression	Normal Kan	20	234.1500	5.0604
	Normal eng	20	223.6500	9.7618
Comprehension	Normal Kan	20	143.0000	4.4129
	Normal eng	20	118.0000	11.4202
TOTAL	Normal Kan	20	463.7000	6.4571
	Normal eng	20	423.0000	18.2064

Table – 2: Mean and S.D of normal individuals in Kannada and English on different tasks.

ii. Comparison of performance of individuals with DAT in Kannada and English Languages

Comparison of performances of individuals with DAT in Kannada and English was done using independent samples t – test to find out the significant difference between the two groups across the parameters. Results showed that there is a significant difference between the performances of DAT individuals in English and Kannada on all the three tasks (p<0.005, f = 0.01). The mean and S.D values showed that the performance of DAT individuals was better in L1 (Kannada) than L2 (English) in all the three tasks. Mean and S.D of both the groups are given in Table – 3.

Tasks	Group	N	Mean	Std. Deviation
Memory	Dem Kan	20	42.6500	5.5562
	Dem eng	20	60.6000	5.1031
Expression	Dem Kan	20	118.0000	11.4202
	Dem eng	20	95.7000	12.2651
Comprehension	Dem Kan	20	98.0500	11.5005
	Dem eng	20	81.3500	4.9553
TOTAL	Dem Kan	20	258.7000	15.3695
	Dem eng	20	237.6500	17.1595

Table – 3: Mean and S.D of individuals with DAT in Kannada and English on different tasks.

iii. Comparison of performance of individuals with DAT with normal individuals in Kannada Language

Comparison of performances of individuals with DAT and normal individuals in Kannada was done using independent samples t – test to find out the significant difference between two groups across the parameters. Results showed that there is a significant difference between the two groups on language expression (p<0.005, f = 0.04) and language comprehension (p<0.005, f = 0.01) tasks. No significant difference was found between two groups in memory tasks (p>0.005, f = 0.06). The

mean and SD values of both the groups are given in Table – 4. The performance of individuals with DAT was very poor when compared to normal individuals.

Tasks	Group	N	Mean	Std. Deviation
Memory	Normal Kannada	20	86.5500	2.4810
	Dementia Kannada	20	42.6500	5.5562
Expression	Normal Kannada	20	234.1500	5.0604
	Dementia Kannada	20	118.0000	11.4202
Comprehension	Normal Kannada	20	143.0000	4.4129
	Dementia Kannada	20	98.0500	11.5005
TOTAL	Normal Kannada	20	463.7000	6.4571
	Dementia Kannada	20	258.7000	15.3695

Table – 4: Mean and S.D of individuals with DAT and normal individuals in Kannada on different tasks.

iv. Comparison of performance of individuals with DAT with normal individuals in English Language

Comparison of performances of individuals with DAT and normal individuals in English language was done using independent sample *t* – test to find out the significant difference between the two groups across the parameters. Results showed that there is a significant difference between the two groups in memory ($p < 0.005$, $f = 0.04$), language expression ($p < 0.005$, $f = 0.02$) and language comprehension ($p < 0.005$, $f = 0.02$) tasks. The mean and SD values are given in Table – 5. The overall performance of normal individuals was better than the individuals with DAT.

| Tasks	Group	N	Mean	Std. Deviation
Memory	Normal Kannada	20	81.3500	4.9553
	Dementia Kannada	20	60.6000	5.1031
Expression	Normal Kannada	20	223.6500	9.7618
	Dementia Kannada	20	95.7000	12.2651
Comprehension	Normal Kannada	20	118.0000	11.4202
	Dementia Kannada	20	81.3500	4.9553
TOTAL	Normal Kannada	20	423.0000	18.2064
	Dementia Kannada	20	237.6500	17.1595

Table – 5: Mean and S.D of individuals with DAT and normal individuals in English on different tasks.

Differences in performances in all the tasks between groups

The mean and standard deviations are compared between groups for each task, i.e., memory, language expression and language comprehension.
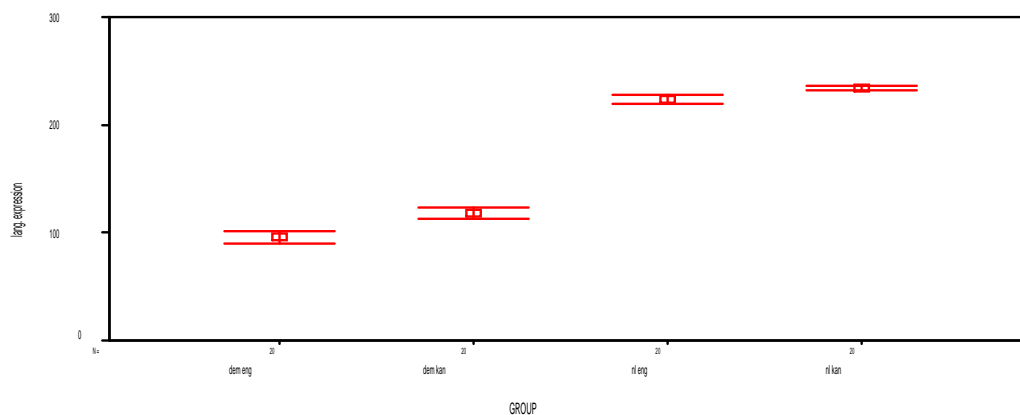
Differences between groups on memory task

Significant difference was found between the four groups on memory task. Normal individuals in significant difference were found between the four groups on memory task. Kannada performed better than other groups and individuals with DAT in Kannada performed poorer than other groups. Comparison of the performance of different groups is represented in graph – 1.
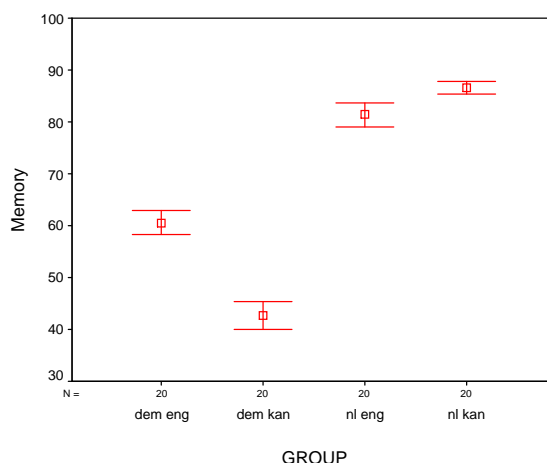
Graph – 1: Comparison of performances of the groups on memory task.

**Differences between groups on language expression task**

Significant difference was found between the four groups on language expression task. The normal group performed equally in Kannada and English and is better than individuals with DAT in English performed poorer than other groups. Comparison of the performance of different groups is represented in graph – 2.

Graph – 2: Comparison of performances of the groups on language expression task.

Differences between groups on language comprehension task

Significant difference was found between the four groups on language comprehension task. Performance of normal group in Kannada was better than other groups and individuals with DAT in English performed poorer than other groups. Comparison of the performance of different groups is represented in graph – 3.
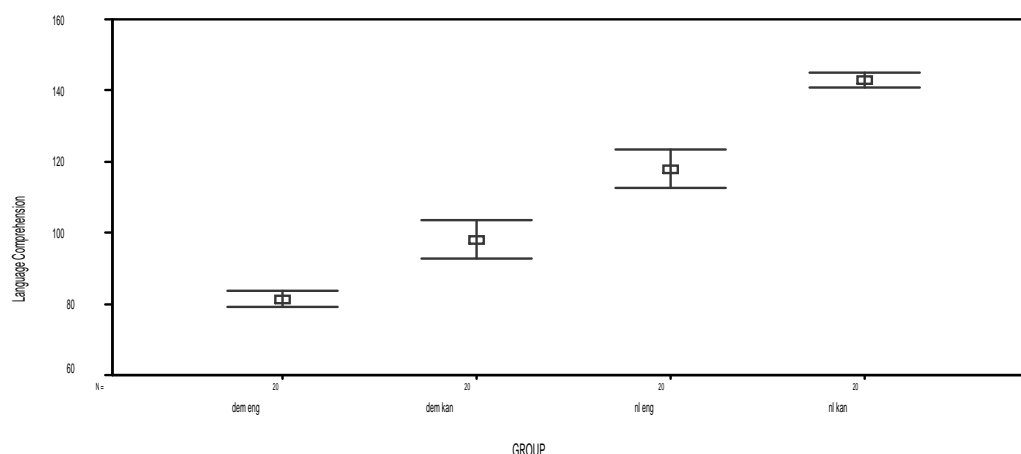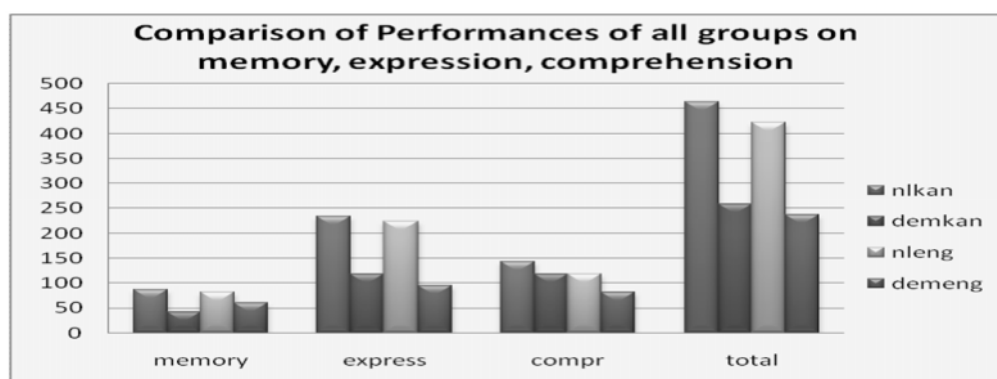
Graph – 3: Comparison of performances of the groups on language comprehension task.

Graph – 4: Results of the all the groups on three tasks and total of the three tasks.

Discussion

The above results indicated significant differences in the performance of language and memory skills between normal and DAT individuals in both languages. The differences in performance between Kannada and English languages in normal group and also in individuals with DAT can be attributed to the factors like age of acquisition of English, amount of exposure, language use in daily life, etc… and these factors can influence the performance of both the groups. This study gives an overview of the language deficits seen in individuals with DAT and these deficits again can depend upon the site of lesion and the premorbid language skills, etc.

Bilingual dementia patients tended towards asymmetrical language impairment with preferential preservation and use of the first acquired language. Studies in aphasic patients from strokes and other brain lesions show that recovering language patterns are most commonly synergistic; recovery in one language is accompanied by recovery in another. Many bilingual aphasic patients, however, recover differentially in one language. In these circumstances, the language most recovered may be the earliest acquired language, or the language of greater use, or the language spoken in the patient's environment. In dementia, recently learned information is retained the least and the older, more remote information is often relatively preserved, consistent with a regression toward the predominant use of the patient's earliest language. In dementia, a retreat to the original language could result from an exacerbation of the cross-language difficulties that typically increase with age. "Cross – language interference" refers to deviations from the language being spoken due to the

involuntary influence of the "deactivated" language. People who are bilingual never totally deactivate either of their two languages, and this can result in interference or intrusions, particularly from the dominant language into the other one. Dementia patients tend to mix languages, and they have special problems with language separation.

6 Conclusions

This study has revealed that the individuals with DAT have language deficits in both comprehension and expression and also in memory tasks. Significant differences were found between the performances of DAT individuals in Kannada and English. This is the case with normal individuals also as seen in this study. This becomes an important issue in the selection of the language used in speech and the language management of bilingual DAT individuals. Individuals with DAT have better preserved the language skills in Kannada as compared to English and thereby, suggesting continuing the speech and language therapy in the native language than in the second language.

The deficits in language skills in both languages were similar to that of bilingual aphasics, but, the recovery processes may not be the same in both the disorders due to variations in the site of lesion and also due to changes in histopathological changes in the central nervous system (CNS). For speech – language pathologists, it becomes an important issue during the selection of management programs for individuals with DAT. Several studies have been conducted on normal bilinguals and on bilingual aphasics to find out the language organization in them. Future studies are expected to determine the language organization in CNS in DAT individuals and their effect on first and second languages. More studies in this area in different languages can give information on differences in performances between languages. These results can be applied and are useful in both the assessment and the management of the individuals with DAT.

References

Bayles, K. A., & Tomoeda, C. K. 1993. *Arizona Battery for Communication Disorders of Dementia.* Tucson, Arizona:

Canyonlands Publishing, Inc.Bayles, K. A., Tomoeda, C. K., Cruz, R. G., & Mahendra, N. 2000. *Communication abilities of individuals with late-stage Alzheimer disease. Alzheimer Disease and Associated Disorders*, 14, 176-181.

Chapman, S. B., Zientz, J., Weiner, M., Rosenberg, R., Frawley, W., & Burns, M. H. 2002. Discourse changes in early Alzheimer's disease, mild cognitive Impairment, and normal aging. *Alzheimer Disease and Associated Disorders,* 16(3), 177-186.

Cummings JL, Benson DF, Hill MA. 1985. Aphasia in dementia of the Alzheimer type. <u>*Neurology*</u>. 35: 394–397.

Cummings, J.L., & Benson, D.F. 1992. *Dementia- A clinical approach.* (2nd ed.). Boston: Butterworth-Heinemann.

De Santi, S., Obler, L. K., Sabo-Abrahamson, H. and Goldberger, J. 1990. Discourse abilities and deficits in multilingual dementia. In Y. Joanette and H. H. Brownell (Eds) *Discourse Ability and Brain Damage: Theoretical and Empirical Perspectives* (New York: Springer), . 224 – 235.

Grosjean, F. 1982. *Life with Two Languages: An Introduction to Bilingualism*, Cambridge, MA: Harvard University Press.

Luderus, L. 1995. *Language Choice and Language Separation in Bilingual Alzheimer Patients.* Amsterdam: IFOTT.

Paradis, M. 1997. The cognitive neuropsychology of bilingualism. In A. M. B. de Groot and J. F. Kroll (Eds) *Tutorials in Bilingualism: Psycholinguistic Perspectives*. New Jersey: Lawrence Erlbaum Associates, Page numbers. 331 – 354.

Shaji. & Bose. 2005. Prevalence of dementia in an urban population in Kerala, India. The British Journal of Psychiatry, 186; : 136-140.

Semantics of the classifiers in some Indian languages

Ritesh Kumar   Bornini Lahiri   Atanu Saha   Sudhanshu Shekhar
riteshkrjnu@gmail.com   lahiri.bornini@gmail.com   atanu.jnu@gmail.com   shekhar921@gmail.com

Centre for Linguistics, Jawaharlal Nehru University. New Delhi

Abstract

In the present paper, we present a detailed description of the classifier systems of five Indian languages-- Mizo, Galo, Tagin (all belongs to the Tibeto-Burman family), Assamese (Indo-Aryan) and Malto (Dravidian). It is observed that the classifiers are a predominant feature in the Tibeto-Burman and we observe an extensive classifier system in these languages. There is no equivalent classifier system in other language families. However in the languages belonging to Eastern India, irrespective of the family, there is some sort of classifier system. Thus classifiers seem to be an areal feature in most of the Eastern and whole of the North-Eastern India. The purpose of the paper is to study if there is some semantic similarity among the classifier systems across language families in this area and thus to see if it is indeed an areal feature. It is just a preliminary description of an ongoing research in which we intend to study many more languages and include languages from the Austro-Asiatic family (such as Khasi and Munda languages spoken in Jharkhand) as well.

1 Introducing Classifiers

There are two kinds of noun classification systems found in the languages of the world— noun class system and classifier system. Classifier languages can again be categorised into six types— noun classifiers, numeral classifiers, classifiers in possessive constructions, verbal classifiers, locative classifiers and deictic classifiers. All these systems classify nouns on the semantic basis. But they differ morphosyntactically as well as on the basis of the preferred semantic features. Noun class systems are those in which the classification of Noun is generally done on the basis of such semantic categories as animacy, sex, humanness and sometimes shape. The noun class systems are marked by a finite and limited number of classes in which nouns can be classified. Thus in a language like Hindi, Nouns are classified into two classes. However, it can go up to 10 as in some Bantu languages and even to several dozen as in some South American languages.

The classifier languages differ from the class languages on the account of always being morphologically represented. It is not necessary for every noun, numeral or verb to occur with a classifier. And one noun may take more than one classifier depending on what property of the noun is being emphasized. In Noun classifier languages, the classifier comes with the Noun while in Numeral classifier languages, it is represented on a numeral or a quantifier (and so it comes only in the cases where noun comes with the numeral or the quantifier). Noun classifiers are generally independent words which typically mark distinctions on the basis of humans, flora, fauna, natural objects, artefacts, food (flesh/vegetable), edibility, portability, kinship, and other culturally important concepts. Numeral classifiers categorize the nouns in terms of its animacy, physical or functional properties, and cultural elements.

2 Galo: A Tibeto-Burman Language

Galo is a Tibeto-Burman language spoken mainly in the southern half of the West Siang district of Arunachal Pradesh and some parts of East Siang and Upper Subansiri also. It is a numeral classifier system i.e., classifiers occur only when a numeral comes with the noun. The classifiers and numerals attach to the noun in the following order— N[NUM-CL+NUM]. Galo classifies the objects according to their shape, size and dimensions. Some of the numeral classifiers used in the language are as follows.

Classifiers	Type of Noun	Example
ho	long slender object	d m ho-ken hair CL-One a strand of hair
p	not very long, but fleshy and is covered	oi p -ken fish CL-One one fish
bor	thin piece of cloth, a leaf or a piece of paper	e i bor-ken clothCL-One a piece of cloth
tak	objects which has hard surface	k ke t k-ken stool CL-One One stool
p	round or circular objects	p p p -ken egg CL-One an egg
pom	bunch of ants, keys or flowers	t r k pom- i Ants CL- two Two ants
ob	measurement with fingers	l kc ob-ken Finger CL-One one finger
ci	a pinch of something	lo ci -ken Salt CL-One a pinch of salt

Besides these numeral classifiers, Galo exhibit a very interesting phenomenon of partially reduplicated numeral classifiers. In cases where nouns do not take any classifier as such, the last syllable of the noun is reduplicated and attached to the numeral. This partially reduplicated syllable functions as the numeral classifier for that noun. Some of the examples are as follows

Classifiers	Noun	Example
l k	Used for hands only	l k l k-ken Hand CL- One One hand
m	for dream, graveyard and boundaries	j m m - m Dream CL-Three

		Three dreams
lo	for day	**lo lo**-ken day CL-One One day

3 Tagin: A close kin of Galo

Tagin is mainly spoken in the Upper Subansiri district of Arunachal Pradesh. It is also a numeral classifier language. It has a very elaborate system of Numeral Classifiers with a small class of basic classifiers, based on shape, size and function. No reference to humanness or animacy is related. Tagin classifiers follow the same pattern as Galo i.e.,N [NUM-CL+NUM]

Like Galo there are some basic classifiers which are based on shape, size and function and there are some classifiers which come for single entity. Some examples of classifiers which map entities based on shape, size and function are as follows.

Classifiers	Type of Noun	Example
ta	two dimensional objects	si tak ta- u board CL-5 Five boards
c	round small objects	t rop c - u ant CL-5 Five ants
so	long, cylindrical objects	g ri so-pi vehicle CL-4 Four vehicles
da	narrow objects	s da- i firewood CL-two
bor	broad objects	e bor- i leaf CL-2 Two leaves
n	reproductive entities	im n - pi woman CL- four Four Women
p	round objects	p p p --kin egg CL--one One Egg
di	certain kind of elevation.	modi di-kin Mountain CL -one One mountain

Moreover Tagin also exhibits the phenomenon of partially (as well as fully) reduplicated classifiers. These classifier morphemes are derived from reduplicating the last syllable of the referent nouns. Here are a few examples:-

Classifiers	Noun	Example
harpen	half pants	harpen pen—ri

		half pant CL —ten ten half pants
ma	dead bodies	s ma ma—om dead body CL –three three dead bodies
kuk	bag	rakuk kuk- pi bag CL -seven seven bags
ta	shoes	juta ta- ri shoe CL-ten ten shoes

4 Mizo: Another Tibeto-Burman Language

Mizo is again a Tibeto-Burman language spoken throughout the state of Mizoram, parts of southern Manipur and the eastern Zampui hills of Tripura.

Mizo is also a classifier language and it has both Noun classifiers and Numeral classifiers. Only animates in the language takes up Noun classifiers, inanimate are not classified. The distinctions for Noun classifiers are mainly made by Gender Markers in [+Human] while the classifiers for [-Human] are Gender Markers and Change of State like pre and post- reproduction as well as pre-nailed/horned or post-nailed/horned. The distinction for Numeral classifiers is made from the shapes and size of the objects. For Noun classifiers, Mizo follows N-CL word order while for numeral classifiers, it follows the usual Tibeto-Burman order of N[NUM CL+NUM].

There has been an interesting observation in Mizo where the numeral classifiers for mass nouns [+cumulative] take a state of action.

Mizo Noun classifier system

The system of categorization of nouns in Mizo reveals the fact that the language is a noun classifier language. The evidences for this proposal are as follows:

It is a fact that in a noun classifier language every noun does not necessarily take a noun classifier and a noun can occur with more than one classifier.

In Mizo [pa] is a classifier used to denote masculinity and it comes with human nouns especially. But it may also occur with professional nouns as given below:

Zirtirtu-pa   zirtirtu-nu
Teacher-CL   *teacher-CL*
Male teacher.   female teacher.

We can also describe the classifiers in terms of the following three features:

●*Obligatory:*

These noun classification devices are sometimes obligatory sometimes not in Mizo and explicit rules can also be formed for their omission. For example in borrowed words classifiers can be dropped.

●*Productivity:*

These are very productive in the system and even if a new noun comes into the language classifier is more or less fixed to it.

●*Independence:*

Like any other classifier language Mizo classifiers can be used as an independent word.

Example: - pa which is a human masculine classifier is also used as the word for father with a different tone.
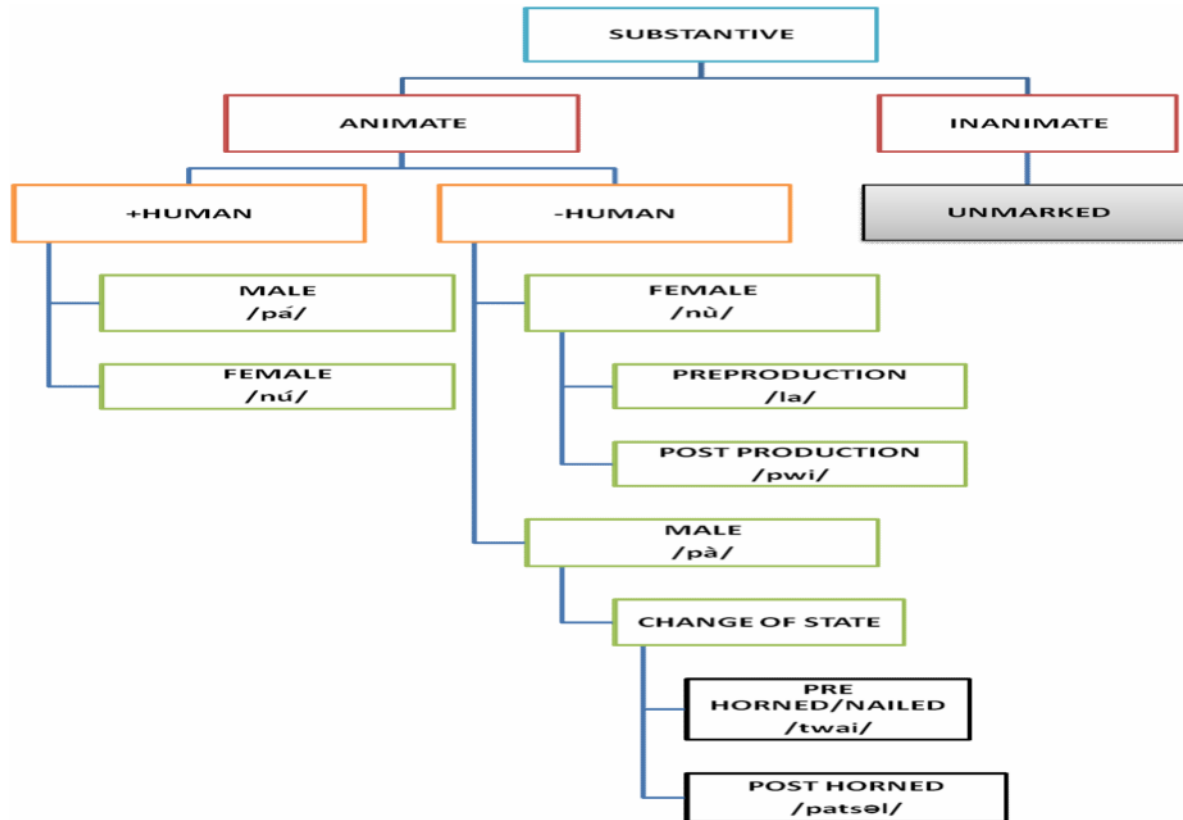

Figure 1Mizo classifier system

Classifier for human male is /pa'/. This is a tonal language and that's why tone distinction is very important. Human Female classifier is /nu`/. /pá/ and /nú/ are also words for father and mother respectively. For the common distinction between male and female animates, classifiers are /pà/ and /nù/ accordingly. Animate classifiers are also distinctive in case of change of state (reproductive stage and coming out of nail). Horned animals like sheep, goat, and cow take the classifier /pat l/ with them. /pat l/ is attached as a classifier to the animals having a nail at the back of leg (cock). Classifiers are different in case of pre-reproduction and post reproduction stages for animates. Before reproduction male animate classifier is /twai/ and female is /la/. After reproduction they are /t l/ and /pwi/ respectively. For proper names there are distinction between male and their female counter parts. Classifier /-a/ is adjoined to male names and /-i/ with a female name.

Mizo numeral classifier system
In Mizo all Human nouns occur with the classifier [mí]. For elongated objects like cigarette, bamboo, wood and battery [tl :n] is used. Round objects come with [púm]. Flat objects such as paper and plywood take [p e:k] as classifier. For tree and flower classifiers are [ku:] and [pa:r] respectively. [t i] is used to denote the kind of things i.e. the kind of stories, etc.Pieces of objects take the classifier [t ém]. For fruits, egg yolk, eye ball , orange seeds, etc. [mú] is used as the classifier. Food grains are referred by the classifier [fa:]. [zai] is used to denote small elongated objects like hair, wire , and thread.
5 Assamese: A neighboring Indo-Aryan Language

Assamese, which belongs to the Indo-Aryan language family, has a wide range of Noun and Numeral classifiers. However, the number of Numeral classifiers is larger than that of Noun classifiers. Most of the Noun classifier move to a slot near the numeral in the NP and become Numeral classifier. However, there also remain some classifiers which do not raise and become a numeral classifier. The different Noun classifiers can be distinguished on the basis of some semantic concepts such as shape, size, structure, +/-humanness, +/-animacy, +/-honorific, etc.

Some of the classifiers in Assamese are as follows

| Classifiers	Type of Noun	**Example**
zupa	plants and trees	aam-zupa mango tree-CL a mango tree
k n	flat structured objects	bis na-k n bed-CL a bed
dal	branch like/stick like long objects	suli-dal hair-CL hairs
k ila	leaves and pages	pat-k ila leaf-CL a leaf
pah	flowers	gulap-pah rose-CL a rose
pat	flat structured object that is worn	kan-p uli-pat Ear-ring-CL a ear ring
t upa	bunch of flowers	p ul-t upa flower-CL flowers
tupal	liquid drops	pani-tupal water-CL water

6 Malto: A neighboring Dravidian Language

Malto belongs to the Dravidian family of languages. In this language objects are classified along the lines of discreteness and Non discreteness. Animates are classified into +/- Human. Some Classifiers are formed by partial reduplication. It's a very interesting classification system. The general structure of the Noun Phrase is Numeral+ Classifier+ Noun. Animate entities are classified into human and nonhuman.

tini	jan	maler [1]
Three	CL	men

Three men

tini	maq	bu:te[2]
Three	CL	ghosts

Three ghosts

The reduplicated classifiers are applicable to restricted sets of Nouns such as qep- 'village', man- 'tree', ki - small plant, ka i- hole/cave, kuji- 'shadow/reflection', etc. Interestingly when the numbers are one, two and the noun is –Human the structure of noun phrase changes to Classifier+numeral+Noun.

maq-s	go:	ro
classfier	two	horse

two horses
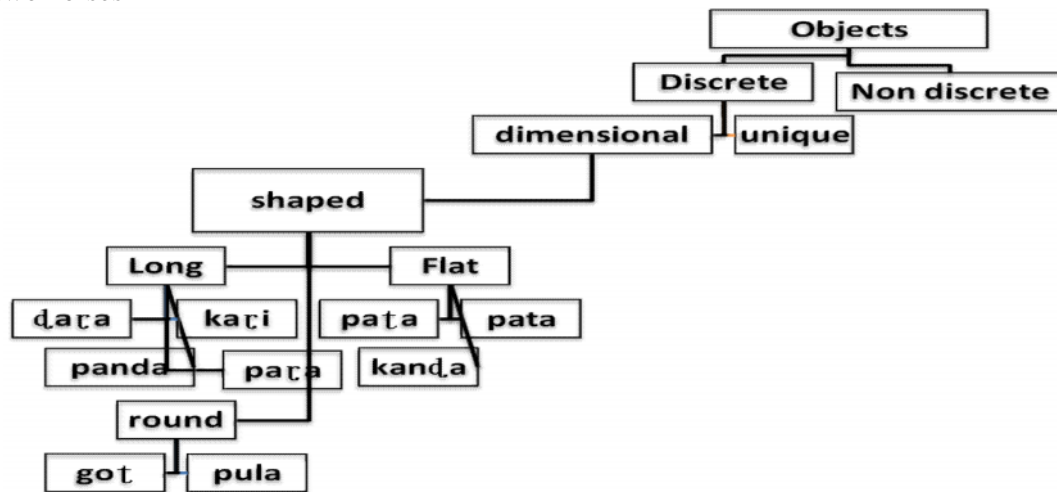
Malto system of Classification.

Some of the state classifiers in Malto are: '*baha*'- Area, field , market ,place, ' a a'-- small area ' ika'– clod , earth.

When the noun is +human portmanteau morphs 'ort-' for one and 'irw-' for two occur respectively. Classifier 'jan' is used with human nouns for live referents. 'maq' comes for their non human counterparts. For objects which are long and large the classifier is ' a a'. 'ka i' comes with long small objects. Objects that are both long and flexible take 'panda' as a classifier. Use of 'pa a' is frequent in long fruits. Malto classifier system also distinguishes objects along the line of flatness. Flat, broad objects are used with the classifier 'pa a' and 'pata' for flat, broad but thin objects. Kan a is however used with cotton objects that are flat and broad. Round heavy objects occur with 'go ' but round, light objects will definitely be produced with the classifier 'pula'.[3]

An interesting construction in the noun phrase can be seen for some nouns e.g. rice, water, etc.

Bu   p ur   k t

---

[1]   Mahapatra, B.P. article on '*Malto Numeral classifiers*' published in *Languages of the tribal and Indigenous people of India.* Abbi, Anvita (ed.) 1997 MLBD SERIES IN LINGUISTICS VOL. X. DELHI: Motilal Banarsidass publishers.

[2]   ibid.

[3]   Pa:n and pan   are other two variant forms of this for the number 1 &2 respectively.

Rice CL-action of NUM (one)
carrying rice in a basket
One traditional basket of rice.

tui   k t   k t
Water   CL:action of carrying NUM (one)
Water in two buckets with a bamboo pole
Two buckets of water

The same phenomenon occurs with the classifiers in mass nouns [+CUMULATIVE] where mass nouns are classified as a state of action. This is interesting because the action predication seems to be attached with the head noun and serves the purpose of a classifier. However it is not the case that main predication can be dropped even if these phrases are used at sentence level. This is true for mass nouns and indeed can be said a strong case of grammaticalization. Reduplication is an important phenomenon in almost all languages and it is shared across language families. Let's make a comparison between Tibeto Burman languages and that of the Dravidian family.

Name of the language family	Language	Example	Remarks
TB	Mizo	Pa: par par k t	Second syllable is reduplicated
	Tagin	harpen pen –ri half pant CL-Ten ten half pants	
	Galo	d m p p - m Pillow CL-Three three pillows	
DA	Malto	qep-ond qep u CL-NUM village One village	First syllable is reduplicated

7 Conclusion

On the basis of the above discussion, it could be seen that there are classifiers in languages belonging to three different families—Tibeto-Burman, Indo-Aryan and Dravidian. While Tibeto-Burman languages have been widely accepted as classifier languages, Dravidian and Indo-Aryan languages are generally not considered so (even though Eastern Indo-Aryan are being slowly acquiring the status as classifier languages). The presence of this feature in these languages indeed shows that classifiers are an areal feature of these languages. Due to the contact with TB languages for a long time, classifiers may have been borrowed into these languages (very tentatively). However as we could see classifiers in all these languages have very unique features and behave in language-specific ways (except Galo and Tagin, which are very closely related). This is true even for Mizo and the other two TB languages. And because of its uniqueness it cannot be established

with conviction whether classifiers have developed by themselves in these languages or they are borrowed. If they are developed by themselves then why have they developed only in these neighbouring languages? And if they are borrowed how come they are so distinct that they bear very little resemblance, if any, with each other? It looks like a more nuanced case of linguistic diffusion where a linguistic feature has been borrowed but the exact characteristics of that feature were developed within the language; it was not just borrowed but also worked upon and developed according to their own needs. This aspect needs to be explored more to arrive at some definite conclusion.

References

Aikhenvald, A Y. *Classifiers: a typology of noun categorization devices*. Oxford: Oxford University Press(2000).

Aikhenvald, A Y .*Classifiers and Noun Classes: Semantics*, from 'Encyclopedia of Language and Linguistics' eds Keith Brown (editor-in-chief), Anne Anderson, Laurie Bauer, Margie Burns, Jim Miller and Graeme Hirst (Elsevier Pvt. Ltd., 2006).

Kidwai, Ayesha and Joyshree Sutradhar (2008). *The Morpho-Syntax of Classifiers in (three) Indian Languages: The Mass/Count Distinction*, presented in CGIML-08 at IIIT-H.

Mahapatra, B.P. (1997). *Malto Numeral classifiers,* from 'Languages of the tribal and Indigenous people of India' ed Anvita Abbi (Motilal Banarsidass, 1997).

Saha, Atanu. *Noun Classification system in Mizo,* from 'Language in India' ed M.S. Thirumalai Vol. 8(2008).

Main verb and light verb in Bangla: Order and intervention

Syed Saurov
School of Language Sciences,
EFL-University, Hyderabad, India

Abstract

The aim of this paper is to carry out a preliminary theoretical and analytical investigation into the syntactic and semantic distinctions between the main verb and the light verb constructions in Bangla. For a native Bangla speaker, there will be no difference between the meaning of sentence (1) and sentence (2) both of which mean 'I have sent the letter' if they are presented as isolated iterations.

> 1)ami   cit^hi   -ta   **pat^hi- e- chi**
> I-NOM   letter DET   send-PRSNT-PERF
> I have sent the letter
> 2) ami   cit^hi -ta   **pat^hi-e diy- e- chi**
> I-NOM   letter DET   send-CP   give-PRSNT-PERF
> I have sent the letter

But Sentences (3) and (4) below show a distinction between the semantic licensing of the light verb and lexical verb in a bi-clausal structure. The main clause means 'I have sent the letter', while the subordinate clause means 'when John came and gave the news'.

> 3)   John   jok^hon eS-e   k^hobor –ta   di-lo,
> John-NOM   when   come-CP newsDET   give-PAST
> When John came and gave the news,
> (totokkhone) ami   cit^hi -ta   **pat^hi- e   diy- e- chi**
> (by then)   I-NOM letter DET   send-CP give-PRSNT-PERF
> I have sent the letter

> 4)   *John   jok^hon eS-e   k^hobor -ta   di-lo,
> John-NOM   when   come-CP   news   DET   give-PAST
> When John came and gave the news,
> (totokkhone) ami   cit^hi -ta   **pat^hi- e -chi**
> (by then)   I-NOM   letter DET   send- PRSNT-PERF
> I have sent the letter

What is interesting is that to express the relation between the meaning of the propositions in the main clause and the subordinate clause, a native Bangla speaker has to use the light verb 'pathi-e   diy-e-chi'. He cannot say 'pathi- e- chi', even though he/she thinks that they are synonymous when his/her judgement is asked about sentence (1) and sentence (2). So there has to be some mechanism which is controlling this, that is allowing only one of the two apparently similar forms. This means there has to be some difference, syntactically and semantically, between these forms. Focus may play an important role here, and the light verbs can contribute to the difference of the sentences. In this paper, I will try to give a possible explanation for this phenomenon in Bangla.

Introduction:

In Bangla light verb constructions, the canonical word order is the main verb followed by the light verb. Also the main verb along with the light verb acts as a constituent resulting in no possible intervention. But interestingly there are deviations from the general order, and even more interesting is the fact that interventions are possible in this changed order. The objective of this

paper is to focus on a few of the light verb constructions within the Minimalist Program framework, and attempt a preliminary syntactic explanation for the structure of the canonical order, the changed order, and the interventions possible.

In a light verb construction, the main verb becomes a conjunctive particle ending in **–e** or **–ye** in these constructions, while it is the light verb which bears all the tense-aspect suffixes. For example:

    1.Tul-e   n-e
       Pick-cp   take-imperative-pres
       Pick (it) up
    2.Pathi-ye   di-ye-che
      send-cp   **give-perf-prsnt /3rd**
       have/has sent the letter

ne and **diyeche** are the light verbs here. In **1** the mood is imperative, tense is present, and this is expressed by **ne**. If it is honorific, instead of **ne**, it would be **nin**. In **2** again everything is marked by the light verb; thus **diyeche** says that it is present perfect, and **3rd** person (singular/plural). Bangla does not have any number agreement; hence there is no distinction between the **3rd** person singular or plural here.

The light verb constructions are ambiguous, depending on how one is reading the light verb. To be more precise, the expression can be read as a main verb followed by a light verb, and also as two separate main verbs. For example, **1** will have two readings as follows:

    3.**Pick up**, where **ne** acts as the light verb, and takes all the affixes though the meaning essentially comes from the main verb **tule** (pick)
    4.**Pick it, and then take** (something), where **tule** and **ne** both act as main verbs, so the meaning is like that of a compound sentence. That is, **tule ne** can be also written as (semantically equivalent of the second reading available):
    5.Tol, ebong ne
      Pick, and take

I have already mentioned earlier that the canonical order in Bangla is the main verb followed by the light verb. The light verb does not have the full verb strength as it has when used as a main verb; for the light verb with the main verb is considered as one single unit (with more strength on the main verb), and any intervention between them renders the sentence bad.

    For example:
    6.*tul-e ekhon n-e
  Pick-cp now take
  *Pick up now

6 is bad because of the intervention **ekhon** between **tule** and **ne**. So the meaning conveyed by the single unit (**tule ne**) is destroyed. However if we consider the second reading **(8)** of **tule ne**, where **tule** and **ne** are read as two independent verbs, the intervention does not render the expression bad.

    7.Tul-e ekhon n-e
      pick-cp now take
      pick up and take something different

7 is perfectly grammatical in the sense 'pick it up, and then take (something) now'.

But the order (the main verb always preceding the light verb) does not hold true in all cases. This is evident from the following expressions:

    8. n-e tul-e
      take pick

pick up

9. di-ye-che   pathi-ye   cithi-ta
   give-perf-prsnt /3rd send-cp letter-det
   have/has sent the letter

In both the cases, the light verbs **ne** and **diyechhe** (respectively), are scrambled before the main verb. Interestingly, here intervention is possible:

10. n-e ekhon tul-e
    take now pick-cp
    pick up now

11. di-ye-che   cithi-ta   pathi-ye
    give-perf-prsnt/3rd letter-det send-cp
    have/has sent the letter

Now to account for the deviation from the canonical order, I will say syntactically the light verb is above the main verb, and it is movement which results in the output order, the aforesaid difference in the order, and also the interventions. It will be clearer with one example:
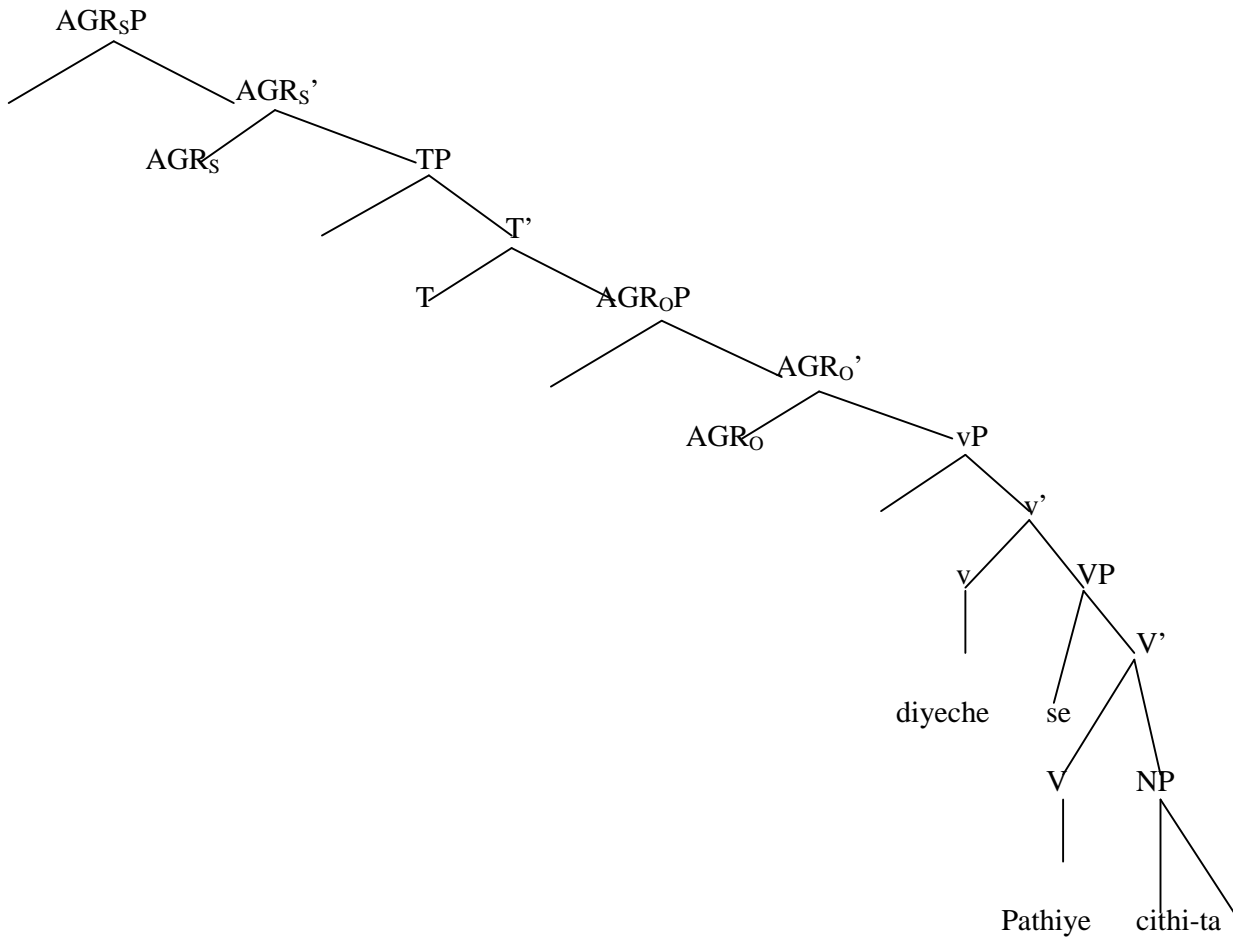
12. Se Cithi-ta pathi-ye di-ye-che
    He letter-det send-cp give-perf-prsnt-3rd
    He has sent the letter

Following Larson's VP- shell model (Larson, 1988), Chomsky, in The Minimalist Program (Chomsky, 1995), says that indeed there is a double VP-shell, but instead of the empty verb position (e) in the outer shell, he proposes that there is a light verb position there. This light verb position is denoted by v, and the outer shell thus being vP.
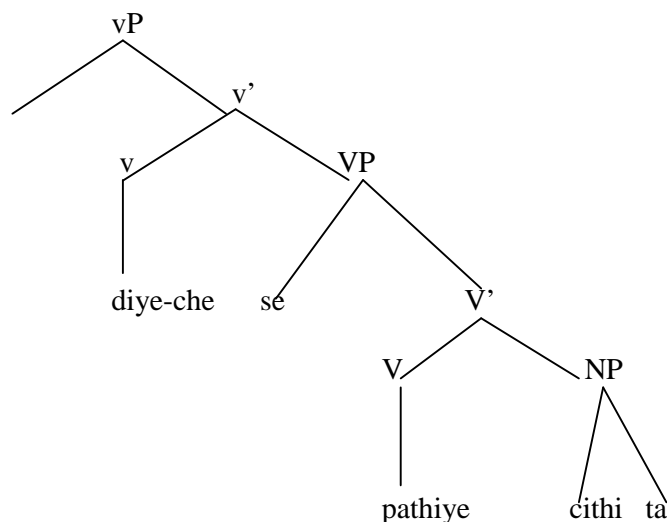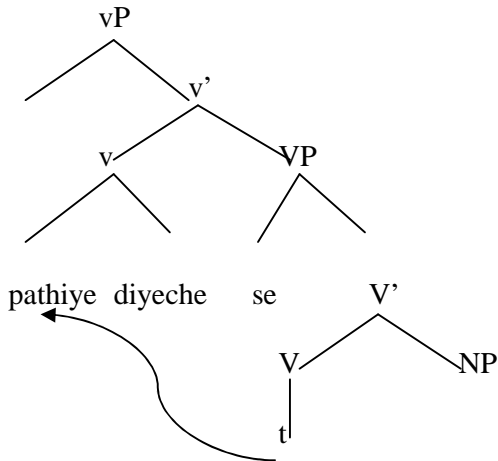

In the Minimalist Program framework, **di-ye-che** and the main verb **pathi-ye** have to check their Tense and Agreement. In minimalism, the earlier proposed IP (Lectures on government and binding, Chomsky 1981) is split into Tense and Agreement phrases, and there are two AGRP (Agreement phrase), one preceding TP, and the other following it. The former one is called AGR$_S$P

(for this is related to the agreement of the Subject), and the latter AGR_OP (as agreement of the Object happens here). Now keeping this structure in mind, **12** can be structurally represented as:

```
AGR_SP
  \
   AGR_S'
   /    \
AGR_S    TP
        /  \
            T'
           /  \
          T    AGR_OP
              /    \
                   AGR_O'
                   /     \
               AGR_O      vP
                         /  \
                             v'
                            /  \
                           v    VP
                           |    / \
                      diyeche se   V'
                                  /  \
                                 V    NP
                                 |    / \
                            Pathiye  cithi-ta
```

The light verb in the structure precedes the main, but as we have seen earlier, apart from certain constructions (**8** & **9**), the order of main verb preceding the light one is always fixed. So the main verb moves and adjoins to the light verb, rendering the word order right (the following tree explicates it).\

```
        vP
       /  \
          v'
         /  \
        v    VP
       / \   / \
pathiye diyeche se  V'
                   /  \
                  V    NP
                  |
                  t
```

The whole complex structure [$_v$ V v] then raises to AGR$_O$ forming the complex [$_{AGRO}$ [$_v$ V v] AGR$_O$], which in turn raises and adjoins to T forming yet another complex [$_T$ [$_{AGRO}$ [v V v] AGR$_O$] T]. Here [$_v$ V v] checks Tense, and then this whole complex raises and adjoins to AGR$_S$ forming [$_{AGRS}$ [$_T$ [$_{AGRO}$ [$_v$ V v] AGR$_O$] T] AGR$_S$]]. Here the [$_v$ V v] checks its agreement with the Subject which has moved successive cyclically to the SPEC$_{AGRS}$. Here as the main verb raises and adjoins to the light verb to form [$_v$ V v], the whole complex becomes a constituent. And hence, intervention here between the main verb and the light verb is not possible.

But what happens in those cases where the light verb precedes the main verb in the output (for example, **9: diyeche pathiye cithi-ta**)? I will argue that the structure proposed for **12** is what the underlying structure is. That is, **9** starts with the same structure with the light verb having a scope over the main in the tree. But here the light verb is being topicalised, or that's what I think for there is no new information available to make it focused. So the topicalised light verb, following Rizzi (1997) where he splits the CP domain into ForceP, TopP, FocP, TopP, and FinP, raises to TopP. Thus the main verb doesn't have to move, whereas the light verb does; and it is the light verb that has to check the Tense and Agreement. So while moving to the TopP, it first goes to AGR$_O$ forming [$_{AGRO}$[v AGR$_O$]], then that complex rises to T forming [$_T$ [$_{AGRO}$[v AGR$_O$]] T] where v checks its Tense. Then the whole complex moves to AGR$_S$ forming [$_{AGRS}$ [$_T$[$_{AGRO}$[v AGR$_O$]] T] AGR$_S$] where v checks the Agreement with the Subject in a Spec-Head relation. Finally the complex structure is raised to TopP. Hence, the word order of the output (**diyeche pathiye,** that is, the light verb before the main) is maintained.

Now as noticed earlier, in **11**, intervention is possible without rendering the sentence bad. Let us consider that sentence again, and see how this can be explained.

11) di-ye-che   cithi-ta   pathi-ye
     give-perf-prsnt /3rd letter-det   send-cp
     have/has sent the letter

The Object **cithi-ta** has to move to the SPEC$_{AGRO}$ where it checks its case with AGR$_O$ in a SPEC-HEAD relation. And the light v **diyeche** has already moved to the FocP, so the word order we get is **diyeche cithi-ta pathiye**, which is the same as the output in **11**. Thus this intervention can be explained. But **diyeche pathi-ye cithi-ta** is also a grammatical sentence. If **cithi-ta** moves to SPEC$_{AGRO}$ for case checking, then how can this word order be explained? Within the mentioned framework, I don't have a proper explanation for this; though I will say that the raising of the Object **cithi-ta** to the SPEC$_{AGRO}$ is in the covert syntax here. Thus it is explained.

A problem with the analysis:

In **10** and **11**, it is evident how intervention is possible in the changed word order of the main verb and the light verb. But there are certain constructions where intervention is possible in the canonical word order:

13) pathi-ye   to   de   ekhon
    send-cp   give   now
    send now (('to' here gives the meaning that send it now, later you can think about it)

14) pathi-ye-i   dilam
    send-cp   give-simp past-1st
    so I did send (this will be the nearest gloss)

However, the preliminary analysis that I attempted cannot account for these cases. Interestingly the interventions in **13** and **14** (**to** and **–i**) are hard to gloss; for the usage is very pragmatic. They add some kind of emphasis or casualness to the sentences. So one way of making the data fit into

the scheme of things would be to say that syntactically **13** and **14** are not any different from **15** and **16**.

15) pathi-ye   de   ekhon
     send-cp   give   now
     send now

16) pathi-ye   dilam
     send-cp   give-simp past-1st
     I sent

Thus if it is said that the interventions in **12** and **13** do not occur in the syntactic component, but they occur rather in the PF, then the data can be incorporated in the analysis. It's not a satisfactory explanation though.

To sum up, in this paper I have tried to provide an explanation for the canonical word order of the light verb constructions in Bangla, the possible change in that order, the interventions possible in that changed order, and the possible interventions with no change in order.

References

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

…….. 1995. *The Minimalist Program.* Cambridge, Mass: MIT Press.

Butt, Miriam. 2003. The light verb jungle. *Harvard Working Papers in Linguistics.*

Dasgupta, Probal. *Bangla Vector Verbs and their Selectivities.*

Deoskar, Tejaswini.2006. *Marathi Light Verbs.* Cornell University

Larson, R. 1988. On the double object construction. *Linguistic Inquiry* 19:335-92.

Lightfoot, D and N. Hornstein, eds. 1994. *Verb Movement.* London: CUP.

Pollock, J-Y. 1989. Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry* 20:365-424.

Rizzi, L. 1997. The fine structure of left periphery. In L. Haegemen, ed. *A Handbook of Syntax.* Dordrecht: Kluwer.

Fast mapping in the developing lexicon in kannada speaking children

Ms. Trupthi T.[1]   Ms. Deepthi M.[2]   Ms. Shwetha M.P[3]   Ms. Deepa M .S[4]

Dept. of Speech-Language Pathology,

J.S.S. Institute of Speech and Hearing, Ooty Road, Mysore

[1]rao.trupthi@gmail.com [2]deepthi_snh@yahoo.co.in [3]shwth.jain@gmail.com
[4]deepams12@gmail.com

Abstract

One of the phases of word learning in toddlers is fast mapping. The incidental learning of new vocabulary in few encounters is known as fast mapping. Researchers have found that children are capable of mapping various aspects of a novel word. This study is aimed at investigating the underlying nature of fast mapping by exploring children's emerging ability to access information in lexical memory. Twenty children between the age range of 2- 3 and 3-4 years, divided into two groups were taught 10 unfamiliar objects over the sessions. Learning of new words by fast mapping and retrieval ability were examined. In the fast mapping phase of word learning process the mean number of trials required for the children to fast map the target words were less over the sessions. In the retention trial the mean percentage of words recalled increased over the sessions and the older group performed better than the younger ones. The results indicated that the children in the higher group could fast map and recall more number of words than the lower group. In our study children were provided with limited information about the target words, yet they were united by all contexts in which the words were acquired.

1 Introduction

Learning to talk is a relatively orderly process, although not all children acquire all language ability in precisely the same order and at identical speed. Though there are a few individual variations. Language learning is synergistic in nature. All components of language- syntax, semantics, phonology and pragmatics-interact to evolve gradually into adult like competence.

According to Piaget, children from birth to about 2 years of age are in the sensory motor stage of cognitive development. One of the resulting cognitive achievements is the concept of object permanence, or object constancy. Object constancy is the awareness of objects in the environment even though they may not be immediately visible. It is a basis for the internal representation of environment, mental images, or symbols of those objects and events that exist around the children. These internal representations are related to children's ability to use verbal symbols nothing but, words.

Acquisition of words in children is in terms of both reception as well as expression. In receptive vocabulary, children comprehend first words at about 8 to 9 months of age (Benedict, 1979). At about 13 months of age, children comprehend about 50 words (Benedict, 1979). By 6 years of age their comprehension vocabulary is between 20,000 and 24,000 words, and by 12 years of age it is 50,000 words or more (Owens,2001). The size of a child's vocabulary depends, in part, on the experiences and words to which the child is exposed (Rescorla, Ali, and Christine, 2001), which after the early years leads to considerable variability in vocabulary composition as well as size.

In expressive vocabulary, children express as many as approximately 10 words at the age of 15 months. By 18 months the expressive vocabulary is around 50 words which are increased to 150 words at the age of 20 months. By 2 yrs of age children express 120-300 words and by 3 years of age the size of the vocabulary is increased to 1000 words. Approximately 1600 words are expressed

by children at the age of 4. And by 6 years of age children utter about 2600 to 7000 words. (Dale, Bates, Reznick, & Morrisser (1989): Owens (2001): Reich (1986): Rescorla et al.(2001): Wehrebian(1970): and Zintz(1970)).

There are several ways in which young children are believed to be so good at learning (Rice and Watkins, 1996). One way is the process known as *fast mapping* (Crais, 1992; Dollaghan, 1985; Heibeck and Markman 1987). Dollaghan (1987) describes *fast mapping* as '*a lexical acquisition strategy in which a listener rapidly constructs a representation of an unfamiliar word on the basis of a single exposure to it. This initial representation might contain information on semantic, phonological, or syntactic characteristics of the new lexical item, as well as non-linguistic information related to the situation in which it is encountered.'*

The first meaning may or may not be complete or accurate. It however creates a basis for further refinement as additional experiences with the word in context occur. Children seem to be able to fast map meaning by having only 'incidental' exposures to new words. That is, new words occur in context in a child's ambient environment, and the child is able to discern what the new word means. This is referred to as *quick incidental learning* (QUIL) (Rice, 1990; Rice Huston, Truglio, & Wright, 1990; Rice and Woodsmall, 1988). Any new word and its partial meanings need to be remembered and over time when new contexts are encountered in which a child is exposed to the word, refinements in the meaning have to be made. These authors comment that the process can take weeks, months, or years, depending on the semantic complexity and frequency of the word to be learned.

Fast mapping is regarded as the initial phase in the more extended process of lexical acquisition. According to Carey (1978), only a small portion of information concerning the novel word is mapped into the long term memory. This initial representation may be incomplete or inaccurate (Dollaghan, 1987). Several more encounters are required for mapping the remaining information, in order to refine the initial representation. Although little is stored on the long term basis, fast mapping does facilitate the later lexical acquisition. It is because once the representation is made, it can be assessed and updated. This initial representation makes the subsequent word recognition more quickly (McClelland & Rumelhart, 1986). Over repeated experiences with the word, full knowledge of its meaning will be achieved. If children require "*a number of exposures to a new word or a significant amount of information about it in order to create a new lexical representation for it then vocabulary expansion would likely be a much more protracted process*" (Dollaghan, 1985).

Carey and Bartlett (1978) first documented the phenomenon of first mapping in a pioneering study of pre school- aged children. They observed that 3-year-olds appeared to learn the meaning of a word when contrasted with another word known to the child. For example, when told "bring me the chromium tray; not the blue one, the chromium one," children successfully mapped the unfamiliar color term in the presence of the familiar one. Carey (1987) proposed that children learn the meaning of a word in two separate phases: (a) a fast mapping phase, in which the child establishes an initial link between a word and its referent, and (b) a subsequent, slow mapping phase. In the fast mapping phase, a child has only a partial knowledge of the meaning of the word, whereas in the second phase of acquisition, the initial word representation is supplemented through additional experience, eventually coming to resemble the adult meaning.

Building on Carey's (1987) original work, numerous studies have focused on possible constraints or biases that guide children to interpret words in particular ways. In the experiments, children were typically exposed to nonsensive learning in an ambiguous context (Gollinkoff, Mervis, and Hirsh-Pasek, 1994; Heibeck and Markman 1987). For example, Evey and Merriman (1998) presented 24-month-olds with two simple line drawings of a cake and a novel object and told them to "find the

dax". Within this context, fast mapping was interpreted as occurring when children attached the new word to the previously unnamed object.

Figure 1

"Point at the dax."

Other studies of fast mapping have focused instead on the speed or robustness with which the initial word –to –referent mapping occurs. Here, the interest is not how children disambiguate the meaning of words but, rather, how general processes of learning and memory support lexical acquisition- for example the number of exposures it takes to learn a new word (Houston-Price, Plunkett, and Haris, 2005) or the kind of information learned and retained in memory (Chapman et al., 1990; Markson and Bloom 1997).

Numerous studies have examined the fast mapping phenomenon in preschoolers in the past decade. Researchers like Carey and Bartlett (1978), Schwartz and Leonard (1984), Dollaghan (1985, 1987), Ross, Nelson, and Wetstone (1986), Heibeck and Markman (1987), Bates, Bretherton, and Snyder (1988) found that children are capable of mapping various aspects of a novel word. It includes its referent, color, texture, function, semantic category, location, action performed on referent as well as its phonological and syntactic characteristics.

All these reflect that children do comprehend the word after the initial mapping stage, rather than just recognize the word. Rice, Buhr, and Nemeth (1990) were interested at revealing children's ability to map different types of words. The results suggested that children map objects and attribute terms better than action and affective words. In addition, the fast mapping ability exhibited by the normally developing and language-impaired children has been recently compared by Dollaghan (1987); Chapman, Bird, and Schwartz (1990); Rice et al. (1990), Rice, Buhr, and Oetting (1992). The findings are inconclusive, because of the difference in the methodology used.

2 Aim of the study

The aim of the present study is to shed light on the underlying nature of fast mapping by exploring children's emerging ability to access information in lexical memory.

3 Method

3.1 Participants

Twenty native Kannada speaking children in the age range of 2 to 4 years (6 boys and 14 girls) were included in the study. All the children were students of Shishya play school situated in Siddhartha Nagar in Mysore. They were divided into two groups based upon their age (2 to 3 and 3

to 4 age). The first group participated in two word learning and the second group participated in three word learning in each trial.

3.2 Stimulus

Thirty brightly colored photograph of real- world object served as stimulus for the study. Target words were selected on the basis of their low incidence in the receptive and expressive vocabulary of young children and were assumed to be unfamiliar to children below the age of 4 years. The words consisted of one and two syllables. Twenty common real world objects were selected randomly from the books of kindergarten. Out of them 10 pictures are used as target ones which are novel words.

A pilot study on 20 children of both the age group was conducted to check the familiarity of the common words and unfamiliarity of the novel words. Result revealed that the target words were unfamiliar and the common objects were familiar to the children.

3.3 Procedure

Children participated in both the training and the testing phase at each session. Each child was shown Microsoft power point slides which consisted of 3 pictures in each slide (2 common and one target). In the training phase the experimenter labeled each of the three target pictures for minimum of 4 to 6 times per session. Naming occurred in an ostensive context when both the child and the experimenter were jointly focused on the referent. The interaction was structured naturalistically. Words were typically embedded in propositional statement to support the processing of word referent pairings. Thus, the child might hear the experimenter say e.g. "idu ha i, ha i mele railu ho:gutade, idu e:nu gotta? Idu ha i"

In each session children were taught novel words. Initially 5 words per session were taken. Since children were not able to learn and recall 5 words in a session, the number was reduced to 2 words for the first group and 3 words for the second group in each session. A total of 5 sessions for the first group and 3 sessions for the second group were required for the children to learn and recall all the target words.

In one session 2 to 3 words were taught and asked to recall after a gap of 10 minutes. Before starting the next session i.e. on the next day children were asked to recall the words taught in the previous session. That served as retention trials and word learning trial started with fresh words. This was repeated for 5 days for the first group and 3 days for second group.

One Microsoft PowerPoint slide consisted of 3 pictures in which one is a picture of the target word and the other 2 are the pictures of the common words. Experimental sessions were ended within 5 days with 5 sessions for the second group and 3 sessions for the first group.

4 Results and discussion

The Graphs 1 and 2 depicts the number of trials required in each session for children in group I and II to fast map a total of 10 target words. The main question was whether children rapidly create lexical representations for the unfamiliar words they encounter. In the fast mapping phase of the word learning process our findings indicated that, in the first group (2 to 3 years; Graph 1) the mean number of trials required for the children to fast map (learn) the target words were less over the sessions. This can be attributed to changes in the strength of lexical activation as a result of repeated practice. Knowing some words appear to prime the system to knowing more words. The more an item, is selected for comprehension or production, the stronger the level of activation will be and hence, the greater the probability of access. Hence there was a decrease in the number of trials across the sessions. This decrease was almost in equal interval from the 2nd to 5th session. But from

1st to 2nd session there was an increase in the number of trials, this was probably because of the retention as well as learning trials together. As the session was increased children were able to recall and learn word with lesser number of trials.

grp 1 learning
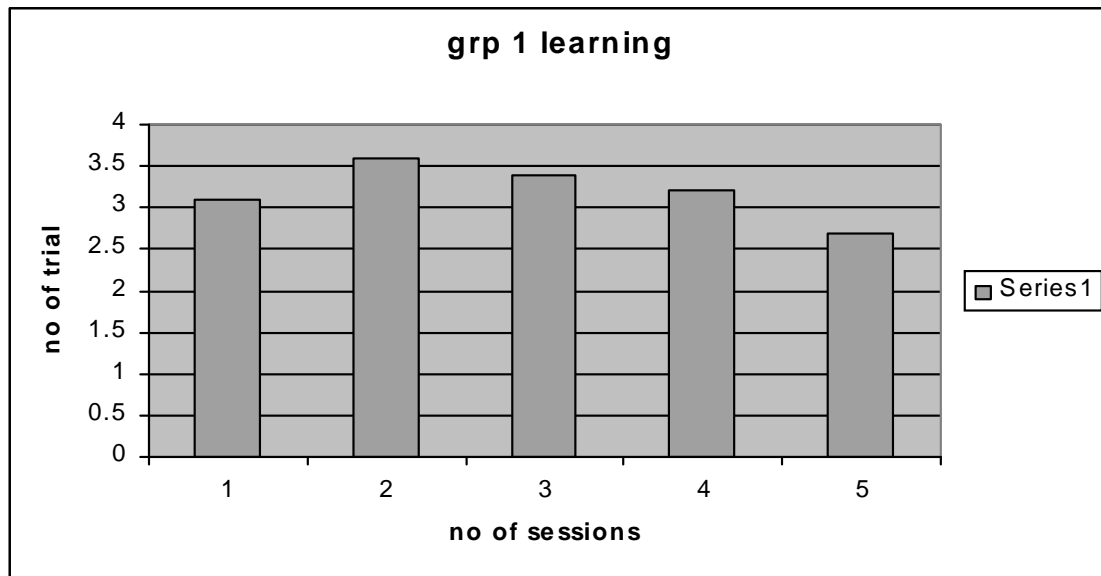
GRAPH 1: Number of trials in each session for children in group I

Graph 2 depicts the number of trials required in each session for children in groupII (3 to 4 years). Children were able to fast map better as the sessions progressed, whereas in the third session the performance deteriorated as the number of words the children were exposed to, was increased from 3 to 4 words. This increase in the number of trial suggest that as children learn more and more new words they require more number of trials. This idea suggests that practice with individual words in a rapidly expanding lexicon changes the operation of the lexicon through the accumulated activation of many items. This is accomplished through an increase in the base of lexical and semantic units and the strengthening of connections between them.

grp2 learning
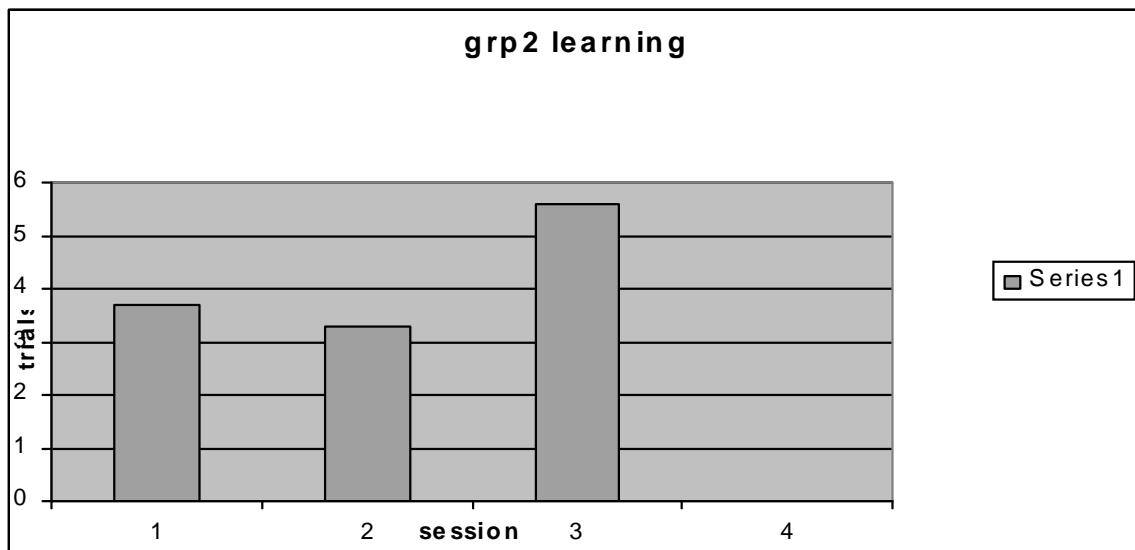
GRAPH 2: Number of trials in each session for children in group II

By comparing the data obtained from group I and group II it is well understood that as the age increases more and more lexical and semantic units can be learnt with fast mapping. The older group learns more number of words in lesser number of trials. The increase in the size of the productive vocabulary increases with age. In the present study we found a modest increase in the receptive and the productive vocabulary in children between 3 to 4 years.

Graph 3 and 4 shows the mean proportion of target words recalled in each session for both the groups. Graph 3 shows mean percentage of words recalled by children of 2 to 3 years.

During the retention trial, number of words children required to recall increased from 1st to 5th session. That is, in each session 2 words were taught and these words along with the words taught in the previous session were to be recalled (1st session= 2 words, 2nd session= 4 words, 5th session= 10 words). As more and more words were fast mapped there was a decrease in the number of words recalled. There was s decrease in the retention of target words in the second session but again the retention increased from 2nd to 3rd session. Whereas the children could not recall as many percentages of words they could do in the first session. The results of the 2nd group is been depicted in Graph 4. Furthermore a total percentage of words learnt by group II was higher than group I with 66% and 45% respectively.
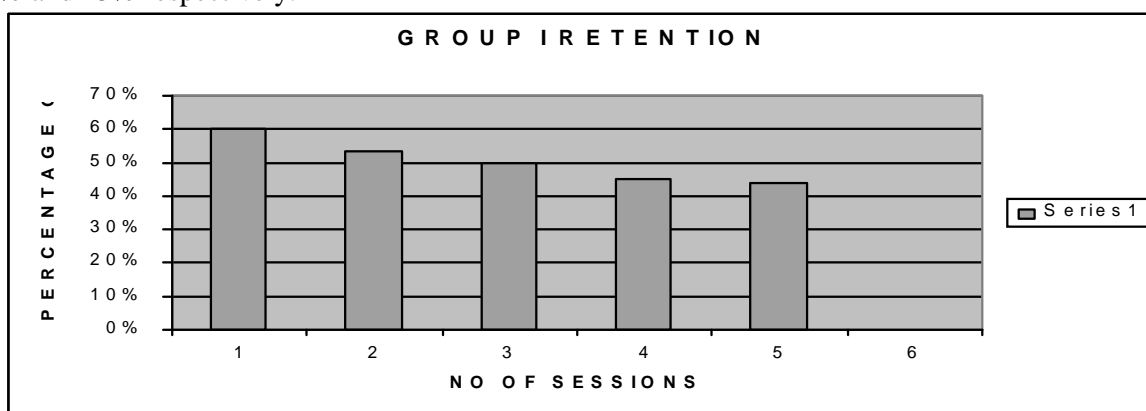
GRAPH 3: Mean percentage of words recalled by children of 2 to 3 years across the sessions.

Additionally there were few observation noted in the retention trial. All children irrespective of the group were able to recall the first target word till the end of experiment (/tapalu/). Children could recall the target words which had more functional value in daily routine for e.g. children could recall /hali/ (railway track) and /sasive/ (mustard) easily compared to words like /hima/ (fog) and /kenda/ (coal). When children were asked to name the target word they were specifying the function of the target word rather than naming it. And again children could recall newly learnt word compared to older ones.

198
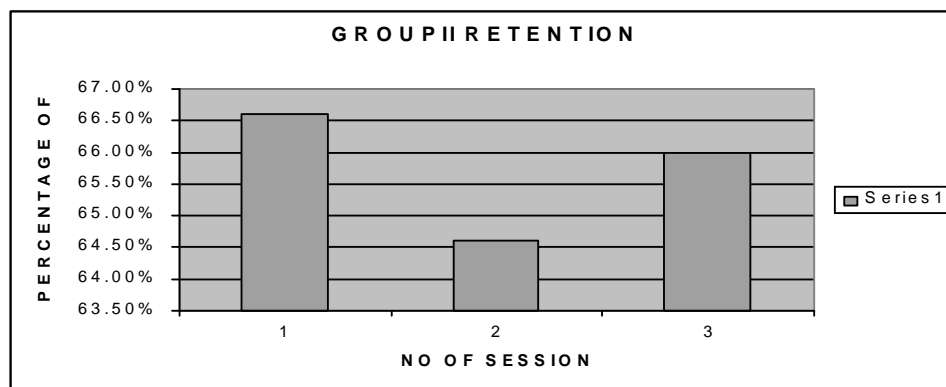
GRAPH 4: Mean percentage of words recalled by children of 3 to 4 years across the sessions.

5 Conclusion:

The present study was aimed at exploring how fast mapping helps in learning lexicon in children. Children selected were between the age group of 2 to 3 years and 3 to 4 years. A set of 10 novel words were taught to them and word learning and retention were examined. All the target words were taught to both groups but the number of sessions required varied. The study opens up many channels in spite of a few limitations. The incidental learning of new vocabulary in the context of one to a few encounters is known as fast mapping (Carey & Bartlett, 1978), or quick incidental learning (Rice 1990). The results indicated that the children in the higher age group could fast map and recall more number of words than the lower age group. Associative learning processes and prior experience, words and concepts can organize into larger units or systems that facilitate fast efficient and parallel access to a broad base of related past knowledge. The idea is that a small set of contextually related words can create structural neighborhood that facilitates the spread of activation to other newly acquired words. In our study children were provided with limited information about the nature and the function of target words, yet they were united by all contexts in which the words were acquired. However the limitation is the number of children taken for the study was less. Gender was not considered as a factor for the study. Standard test for vocabulary development could not be administered. For the 2nd group 4 words were taken for the 3rd session as compared to the previous sessions. And only 10 target words were considered for the study.

References

Bates, E., & Goodman, J. C.(1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and cognition processes, 12,* 507-584.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and reports in child language development, 15,* 17-29.

Chapman, R.S., Kay-Raining Bird, E., & Schwartz, S.E.(1990). Fast mapping of words in event contexts by children with Down syndrome. *Journal of speech and hearing disorders, 55,* 761-77.

Dale, P.S, & Fenson, L. (1996). Lexical development norms for young children. *Behavior research methods, instruments and computers, 28,* 125-127.

Dollaghan, C. A. (1985). Child meets word: "Fast mapping" in pre school children. *Journal of speech and hearing, 28,* 449-454.

Dollaghan, C. A (1987). Fast mapping in normal and language impaired children. *Journal of speech and hearing disorders, 52,* 218-222.

Gershkoff-Stowe, L. (2007). Fast mapping skills in developing lexicon. *Journal of speech, language and hearing research, 50,* 682-697.

Golinkoff, R.M., Mervis, C.B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of child language, 21,* 125-155.

Heibeck, T., & Markman, E. (1987). Word learning in children: an examination of fast mapping. *Child development, 58,* 1021-1034.

McDuffie, A. S. Sindberg H.A. Hesketh, L., Chapman, R.(2007). Use of speaker intent and grammatical cues in fast mapping by adolescents with Down syndrome. *Journal of speech language and hearing research, 50,* 1546-1561.

Reznick, J.S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental psychology, 28*, 406-413.

APPENDIX A: Slides used as stimuli to elicit fast mapping skills.

Beyond Honorificity: Analysis of Hindi *jii*

Gayetri Thakur
Department of Linguistics, Banaras Hindu University, Varanasi, INDIA
gayetrithakur@email.com

Abstract

The present work is about the distribution of a Hindi particle jii traditionally categorized as an honorificity marker or more recently suggested as a nominal classifier. The primary aim of the paper is to provide the different pragmatic usages of this particle along with its mapping in English. The data reveal that jii is used for providing different kinds of speech-act information like seeking permission, requesting, questioning and answering to a question. It can also be used for expressing the mood and the attitude of the speaker towards the addressee. In a conversation (especially telephonic), it is used as a filler by one participant to be with the other participant. After identifying and characterizing the uses of jii, an attempt has been made in the paper to provide the mapping rules for them to disambiguate each one. For instance, in example (3), when the particle is used for expressing the speaker's informal attitude to the addressee, jii is always preceded by kyuN, which can be used as a cue in the mapping rule.

1 Introduction

Politeness is an aspect of language use as important to Hindi speaking community as to any other speech community. There are several devices that make it possible to express politeness in interaction. Hindi makes a distinction in second person pronoun between honorific and non-honorific forms (i.e. *ap* 'you'hon) and (*tum* /*tu:* 'you', non-honorific). Additionally plural forms of third person pronoun (i.e ,*ye*/*ve* 'they'), agreement pattern and certain verbs (*farmaanaa* 'to speak', *biraajnaa* 'to seat') are also used to indicate honorificity.

Politeness, however, should not be considered as an inherent property of the honorific lexical item and syntactic patterns. What is polite in any particular interaction depends on the context of the situation (Firth, 1957; Halliday, 1973). The use of honorific system may express politeness as well as the opposite.

There are a number of linguistic items in Hindi which have no specific grammatical function. However, there functions are innumerable and these elements are termed as particles (avayay or indeclinables). They function as Attitude markers, Fillers, Hesitation markers, Requesting particles, questioning particles, prompters and are also helpful in agreement. They primarily serve various procedural functions though some of them have conceptual meaning as well.

Particles are the elements which don't inflect and the function of these items changes according to the environment. These characteristics of particles makes them most elusive element to work on. They are the functional elements which don't change their form in different syntactic environment but only their interpretation changes. These elements are called functional elements because they have very less lexical content and instead serve as a link to express grammatical relationship between words and speaker's intention i.e. attitude, mood, etc.

The particle used in the present work is a Discourse particle which is frequently found in the south Asian languages and is primarily a feature of the spoken language. The reason behind calling them so is because DPs are those elements that function as signposts in the communication, facilitating the hearer's interpretation of the utterance on the basis of various contextual clues (Aijmer, 2002).

2 Purpose of the study

The work is about the distribution of Hindi particle *jii* traditionally Categorized as an Honorificity marker.

The primary aim of the work is to provide the different usages of the particle along with its mapping rule in English. The particle along with providing different kinds of speech-act information like seeking permission, requesting, questioning and answering to a question in pragmatics also contributes to Lexical Semantics and syntax.

After identifying and characterizing the uses of *jii* an attempt has been made to provide the mapping rules for them to disambiguate each one from the other and this adds relevant theoretical input to Hindi NLP especially in the area of word sense disambiguation for Hindi and Hindi-English Machine Translation.

Although it is difficult to identify the exact syntactic environment that determines the categorical status of this particle in its various uses, in the present work an attempt has been made to list some of the factors that help in identifying the different categorical status of this particle in the language.

3 Methodology

The data has been taken from Hindi magazines, journals, newspapers, etc. and then the different roles performed by the particle *jii* in the language have been explained. This then is assessed on the basis of the Theoretical framework called Discourse markers or Discourse particles (DPs).

The present work reflects the Action structure (term given by Schifrin) of the particle that provides the sequence of Speech-act which occur within discourse. Particles poses complexity in identification and categorization and exhibit difficulty at various levels of processing, they have multiple functional roles based on different syntactic and contextual factors.

4 Different uses of the particle

Different uses of *jii*
A. In Agreement: (in a conversation)
1. Speaker A: *main   kal   d   illii   jaa rahaa huuN.*
                   I   tomorrow   Delhi go-aux-prog-be-1p.sg
                'I am going to Delhi tomorrow'
    Speaker B: *Jii...*
                    Agrp
                    'Right,(sir).'
    7 Speaker A: *tum apnaa kaam puraa kar lenaa*
                    you your work complete do take-infinitive
                  '(You) complete your work'
    Speaker B: *Jii...*
                    Agrp
                    'Okay (sir).'
    8 Speaker A: *paDhaii Suru kii hai naa*
                    study start do aux particle
                        'Did you start (your) study or not?'
    Speaker B: *Jii*
                    Agrp

'Yes (sir)'.

In the given conversations above, *jii* is used as a particle by the speaker in order to agree with the one he/she is interacting with. This occurrence of *jii* is mostly found in formal situations. Hindi provides many morphemes to express one's agreement with the speaker. For instance, *acchaa* 'okay, good' *Thiik* 'right', *haaN* 'yes'. The Honorific marker *jii* while expressing hearer's reverence for the speaker also speaks in agreement with him. Thus, it performs a dual function of expressing honour and carrying agreement simultaneously.

The utterances of speaker B in the examples 1, 2 and 3 can also be responded as

4(a. *accha jii* or *jii accha* 'Alright Sir'

4(b). *Thiik hai jii* or *jii Thiik hai* 'O.K. Sir'

4(c). *haaN jii* or *jii haaN* 'Yes Sir'.

Jii as a bi-morphemic utterance can only be used with yes-no questions when the speaker intends to express his/her agreement with the speaker. A wh-question cannot be responded using *jii* as a complete response. If the hearer does not agree with the speaker he/she has to employ the negative marker *nahii* or *naa* to make his/her point.

B. Requesting Particle

5. *jii aap bhii kuch lijiye*

   Reqp you (hon) also something take-hon-2p

   '(You also) please take something'

In the given structure *jii* occurs in the initial position and is used as a Requesting particle. A Hindi speaker uses this particle in both formal and familiar situation. It is mapped with 'please' in English.

C. In Addressal

6. *aap bhii kuch lijiye, jii.*

   you(hon)also something take-hon-2p-AddP

   'You also take something'

The same particle conveys a different meaning when it occurs in sentence final position as a marker to address the one, the speaker is interacting with and the sentence will occur with a rising pitch intonation in the given situation. This use is found in a familiar situation or just to interact with the participant.

Addressing particle is also used in the following way:

   jii,   suniye

   you(hon) listen

   'Listen or Hey! Listen'.

The occurrence of *jii* in the given sentence is used to address someone when addressing by name is a social taboo. It appears to show politeness towards the addressee and is mostly used by the female speakers (e.g. wife to husband). This can be mapped onto English 'hey'.

D. Vocatives

There is another context where the employment of *jii* is quite common when one person calls another e.g, a teacher calling his/her student or a parent calling his/her son/daughter, etc.

   Teacher: *siitaa*

   　　　　'Sita!'

   Student: *jii sar.*

   　　　　'Yes,Sir.'

In the given conversation *jii* is not used to show any agreement but is just employed in order to respond with 'yes'.

E. Questioning Particle:

9. Speaker A: *vah phir se aa raha hai.*

              he again PP come be-aux-prog

              'He is coming again/back'

Speaker B**:**   *jii?*

              What?

jii in the above context can be used as a question which will be indicated by a change in intonation. (It can also show simple agreement with the speaker of the previous utterance.) This usage of the particle can occur with different intonation patterns, the rising or falling pitch. If the particle occurs with the rising pitch it would mean 'what' and if with a falling pitch it would mean 'ok'. Hence, the intonation pattern plays an important role to disambiguate these two different uses of *jii*.

F. Reinforcing particle:

16. Speaker A: *usne vahaan kyaa kisii ko dekhaa thaa?*

              He   there PTCL anybody ACC saw be pst

              'Did he see anybody there?'

    Speaker B: *jii*

              What/sorry

In the given situation, the hearer being unable to comprehend may respond by uttering *jii* in order to request the speaker A to repeat what he/she might have said. The mapping for the present particle in English will be 'sorry' asking in order to repeat again or 'I beg your pardon'. The speaker can also use 'What' not in interrogation but to repeat the utterances.

The QP and RP can overlap but can be identified with an Interrogative marker in written text and in Discourse it can be identified by the intonation patterns.

G. Attitude Marker

10. *KyuN jii kahaa the itne din.*

    Why part where be-pst so many days

    'So, where were you for so many days?'

The particle is used with *Kyun* in the given sentence in order to mark the informal mood of the speaker.

We frequently use *jii* followed by *KyuN* or the bare particle *jii* in order to begin a conversation in Hindi. It functions as a sentence starter where the interlocutor may use 'so', 'well' in order to begin the conversation.

H. Permission/Hesitation Marker

11. *jii hameN bahar jaanaa thaa*

    PerP we-dat out to go be-pst

    'Actually we wanted to go out.'

The occurrence of *jii* here is in order to seek permission in a formal situation where the speaker is hesitant and is afraid of the reaction of the one whom he addresses. This *jii* is mapped with 'actually' in English.

I. Deferring Particle

12. Speaker A: *tumhaarii fail puurii ho jaaegii to aaj merii madad kar denaa.*

Your file complete be go-fut part today my help do give-inf

'If you complete your file today then help me too.'

Speaker B: *jii …jii*

Ok.. Ok.

In the sentence above, jii occurs as a deferring particle where the 2nd person defers the conversation using jii or jii …jii i.e.o.k along with the sense of respect being expressed.

J. Signalling Sarcasm

13. Mother to child:

sriimaan raam kumaar jii, aap ab paDh leNge?

Mr.   ram kumar hon. you now study take-fut-pl

'Mr.Ram Kumar, will you study now?'

In the given sentence the particle *jii* occurs after the proper name and expresses irritation on the part of the mother. She is using the honorific form sarcastically with the child.

K. Indicating Irritation

14. *Kyaa huaa jii, kyuN nahii paDh rahe ho?*

    What be-pst Irr-Mar why not study be-aux-2p-sg

    'What's wrong? Why are you not studying?'

The above form of *jii* indicates irritation along with the filler *kya* in a situation where the person expresses his/her irritation towards the child. It does not have an exact mapping in English but we often use 'what's wrong' with a little shade of anger instead.

  L. Boost up Marker

15. *jii, aap   sab yah kaam kar sakte haiN.*

    BoostP you (hon) all this work do can be-prt-2p-pl

    'Yes, you all can do this work.'

In the given structure *jii* is used to boost up the participants when produced with emphasis and conveys the sense of 'yes' in English.

5 Syntactic distribution of the particle

Agr p_1a:

It is quite difficult to provide the exact syntactic distribution of the occurrence of the particle. The environment is roughly provided below:

In the first sentence the meaning encoded by the particle is 'Right'.

    4.The occurrence of the particle is determined by the presence of of the Indicative mood of the sentence expressed by the verbal root followed by the aspect and tense marker.

In the second utterance the mapping employed by the particle as a response in English is 'Okay'.

    (a)The particle occurs as a response to a Directive sentence where the transitive verb is in imperative mood and the implied Subject (you) is generally omitted.

In the last sentence the particle is used to respond with 'yes'.

(a) The negative particle '*naa*' used for forming Tag question in the following sentence is a clue which helps in comprehending the occurrence of the particle where the co-participant responds with 'yes'. Such questions are leading questions with a strong expectation of Agreement

  Req p_1b:

(a) Occurs in sentence initial position.

(b) Occurs in a directive sentence (with transitive verb) with overt (or pro dropped) pronoun occurring as the subject of the verb.

(c) The imperative mood of the verb (which is an irregular form) can used as a clue to disambiguate it from other sentences.

Add p_1c:

This use of particle can be identified in the given syntactic environment.

(a). The particle occurs in the sentence initial position followed by some other address word and is separated by comma or an exclamation mark.

Voc_1d:

This use of the particle can be identified in the environment where it occurs before a noun (calling name). The particle is separated by a comma from the noun.

QP_1e:

This particle can be identified in the following environment.

(a).The particle occurs as a response in a conversation. The interrogative marker '?' can be used to distinguish QP and other uses of particle.

RP_1f:

The occurrence of this particle is heavily context dependent and can occur in a case where the co-participant is unable to comprehend. Thus it becomes difficult to provide the exact environment of the utterance.

Atd M_1g:

The occurrence of this particle is determined by the given environment.

(a). 'jii' is immediately followed by some Interrogative adverb or Subject Noun.

(b). In this case a subject noun precedes the particle; the exact environment is quite difficult to determine because of the intonation pattern.

(c).It will always occur in the Second position in a sentence.

Per p_1h:

(a). The particle occurs in sentence initial position.

(b).The sentence is an imperative, to-infinitival construction.

(c).The distinguishing element 'jaanaa' i.e a compound verb occurs with the copula 'be'.

Def p_1i:

Def p is quite difficult to identify. It can properly be identified only in the context of its use.

(a). This use of the particle however can be roughly identified when it occurs as a response to the co-participant in a reduplicated form.

SS_1j:

The use of this particle can be identified in the following syntactic environment.

13.When it is optionally followed by a Subject NP (proper noun) and precedes the honorific use of the pronoun 'aap'.

14.Such usage will always have a transitive optative form of the verb.

I I_1k:

The irritation is again quite difficult to identify because of the intonation patterns associated with it. However, in this use, the particle is always placed before the Wh- word. This can be properly identified in the context of its use.

Boost M_1l:

(a). The particle occurs in sentence initial position.

(b).The sentence is an imperative, to-infinitival construction.

Disambiguation Rules of some of the particles:

I have attempted to provide some of the Disambiguation rules which can be to some extent captured by syntax and in order to provide the rules for each of the usages the topic has to be explored further:

Rules:

- •<jii><noun(human)><verb(transitive/Imperative)>=>jii_1b
- •<jii><noun(human)><verb(modal/Imperative)>=>jii_1l
- •<jii><noun(human)><verb(copula/Imperative)=>jii_1h
- •<jii><Interrogative marker(co-participant)>=>jii_1e
- •<Interrogative Adverb/subject NP)><jii>=>jii_1g

6 Conclusion

In the present work I have examined some of the different usages of the particle *jii* from the point of view of its processing in MachineTranslation from Hindi to English. To capture the different mapping patterns of a particle we need to examine a huge corpus and relate it to its respective category.

Although I have tried to locate as many functions of *'jii'* as possible, there may be some other uses that can be explored further. *JII* has certain other function too which are still unexplored. For instance, certain complex predicates like *jii-huzurii karnaa*('to fawn' in Hindi) or *haanjii karnaa*('to fawn' in Marathi) are many such other instances it has entered into the use of Complex predicates.

The work is important both for the purpose of theoretical Eanalysis and the use of particles in various NLP related task such as tagging, annotation of MT, etc. There is immense Semantic and pragmatic value of the work.

In Hindi grammar there hardly is any discussion on the issue. Therefore the paper is an initial attempt to explore it.

One of the best possible ways of sharpening our description and understanding on the study of DPs can be achieved through the contrastive study of DPs in two or more languages.

Acknowledgments

I would like to thank Dr.Sanjukta Ghosh for providing me an exposure to such an interesting topic and solving my queries from time to time. Her able guidance throughout my work was indispensable.

References

Aijmer,Karin 2002. *English Discourse Particles*. Anna-Maria De Cesare: University of Chicago.

Bhat,Raj N 1989 *Pragmatics of ji: in Hindi*. Maharshi Dayanand University.Research Journal (Arts)*,* Vol:-IX

Brown P and S.C Levinson 1987. *Politeness,Some Universals in Language Usage*. Cambridge University Press.

Crystal. D 1969. *Prosody systems and intonation in English*.Cambridge University Press.

Ghosh,Sanjukta 2006.*Honorificity-marking words of Bangla and Hindi:Classifiers or not*.Bhashacintan,Vol:-1

Guru,Kamta Prasad 1925. *Hindi Vyakran*. Nagari Pracharini Sabha Kashi.

Oesman, J.O 1981 *You Know.A Discourse Functional Approach*.John Benjamins.

Kachru,Yamuna 1980. *Aspects of Hindi Syntax*.Manohar,Delhi.

Sinha, RMK and Thakur, Anil 2006. *Disambiguation and Mapping Strategy for selected Post positions in Hindi for Hindi-English MT*.Bhashacintan.A research journal of Linguistics. Vol:-1

A knowledge rich morph analyzer for Marathi derived forms

Ashwini Vaidya
IIIT Hyderabad
ashwini_vaidya@research.iiit.ac.in

Abstract

A comprehensive analysis of morphological forms is important for the computational processing of highly inflectional languages. One of the productive morphological processes is derivation. Traditionally, it implies the phenomenon where affixation changes the meaning as well as the category of the stem. This paper describes a computational approach to derived forms in Marathi.

Dressler et al. (1987) have observed that certain derivational affixes are more productive than others. As productive affixes account for a greater number of forms, a corpus study was carried out to find the most productive ones. This list was compiled along with rules for morphophonemic changes using the open source morph analysis tool Lttoolbox.

As the process of derivation and inflection can operate one after another, the use of nested paradigms was necessary to account for such forms. The paper also describes the merging of the derivational component with existing inflectional paradigms. This enables the recognition of an inflected stem as well as a combination of derivation and inflection. Finally, the results and evaluation of the work along with a discussion of related issues is presented.

1 Introduction

The task of accurate morphological analysis is important for building Machine Translation systems as they provide basic word level information for the system. This information is vital for subsequent processing, viz. Part of Speech tagging, Chunking, etc.

This paper outlines a method for handling Marathi derived forms. As far as derivational morphology is concerned, a derivational affix is traditionally understood as changing the category and meaning of a word. This change in the compositional meaning of a word makes the identification of the root and the suffix important from the point of view of accurate translation. Moreover, certain derivational suffixes are productive (Dressler et. al.) and hence an accurate recognition and analysis is needed in order to have a broad coverage of all the forms.

In Computational Linguistics, there have been several approaches to handling the problem of derived forms, using either rule-based or statistical systems. Rule-based approaches using a large database of linguistic knowledge have the advantage of giving the correct output every time. Moreover, a language which has a great number of forms presents its own problems for statistical learning although there have been efforts made to acquire morphological knowledge for a highly inflectional Indic language like Assamese. (Sharma et al).

This study follows the approach of supplying as much linguistic knowledge to the tool as possible. Hence, it is a knowledge-rich approach. In this paper, the first section describes the main issues for the derivational morphology of Marathi. The second gives an overview of the Lttoolbox kit and the kind of paradigms created. Finally, there is a discussion on the problems and issues raised after a preliminary evaluation.

2 Marathi: Derivational Morphology

There have been efforts in the past to create Marathi morph analysers (Akshar Bharati et al). These morphological analysers handle derived forms to some extent, but it is done mainly through entering the word in the dictionary. Moreover, the morphological process of derivation is followed by inflection and possibly again by derivation. Such complexity has to be modeled by the morphological analyzer.

While building a morphological analyzer from scratch, priority is given to the analysis of inflected forms as they are more numerous in the corpus. The process of derivation on the other hand applies to fewer forms and is more sporadic in nature. For instance, in English almost all nouns in the lexicon will have a plural form but the derivational morpheme *–ness* will attach to fewer forms.

Moreover, unlike inflection derivational processes are idiosyncratic. Affixation takes place depending upon various criteria such as the origin of the root word, category of the root, gender of the root, or certain other semantic property of the root, not always predictable. The form derived after the process of derivation also differs semantically depending on the root word.

At the same time, certain kinds of derivational suffixes are productivesome are more productive than the others. Aronoff (in Spencer and Zwicky eds.) noted that those affixes which do not exhibit allomorphy and which are semantically transparent are more productive and predictable. A preliminary corpus study of derivational affixes for nouns found that this is more or less true. Out of a total of 36,072 words taken from the corpus, it was found that the following suffixes were more in number than others (Note, this was done only for noun-forming suffixes). Also, it may be worth noting that -p a and -ki attach themselves to foreign words readily like 'mædp a' and 'd kt rki '.

Number found in corpus	Affixes
Less than 5	-manə, ikərəṇə, ṭ ʋəkəriː, etc.
Less than 10	-kiː, be-, -yə, -ṭa, etc.
Greater than 20	-pəṇa, -iː, -ika

Frequency of Suffixes

Frequency of suffixes

From this it is clear that at least some forms are more frequent than others and if these are modeled in the analyzer, not only would it capture a linguistic generalization but also make sure that a number of frequently occurring forms in the corpus would be covered. Note that the gerund forming - e is regular enough to be treated as an inflectional suffix as it attaches to almost all nouns.

2.1 Derivational affixes

Beard (in Spencer and Zwicky eds.) describes four different kinds of derivational suffixes. These are *featural*, *functional*, *transpositional* and *expressive*. A typical example of featural derivation is the affixation that changes the natural gender of a word, but does not change its category. Functional, transpositional and expressive affixes are classified on the basis of the change in category and semantics of the base.

The kind of derivational suffixes that have been included in this study are those that lead to a category change as well as some kind of change in the semantics. Affixation that changes gender

for instance is not considered as it is more convenient to make entries in the lexicon for subsequent morphological processing.

The total number of derivational suffixes in the language differs from one grammar to another. An older grammar like Damle's Shastriya Marathi Grammar (234-64) defines many more derivational suffixes for Marathi than the listing given in Pandharipande's Descriptive Grammar (460-517). She lists a little more than 110 derivational suffixes. Of these, the noun, adjective and adverb forming suffixes are more numerous than the verbs. The following table shows the break up of the various suffixes as taken from Pandharipande's Descriptive Grammar.

Affix types	Number
Adjective forming	39
Noun forming	58
Verb forming	2
Adverb forming	15

Derivational suffix by category

The table shows that there are far more adjective and noun forming suffixes than verb or adverb forming ones. In order to further classify them, it is necessary to find the criteria for attachment to a stem. However, this is harder to define. There are a few affixes (about 13 in number) that attach only to Sanskrit or Persian stems. For the others, the stem's property such gender or ordinality is the operative factor. Else, the majority of the affixes cannot be classified on the basis of the stem to which they attach.

2.2 Affix ordering

The process of derivation and inflection is said to occur one after another, according to the Split Morphology Hypothesis, (Stump in Spencer and Zwicky (eds.) 17). This hypothesis states that once the process of inflection takes place, the lexeme will not undergo derivation, hence derivation would always occur first and will be followed by inflection. Although this Hypothesis has been challenged by evidence from other languages, the phenomenon of both kinds of affixes occurring together has to be handled by the morphological analyzer. This process presents its own problems as both phenomena cannot be handled in a similar fashion. If the dictionary being used is knowledge-rich, then rules have to be written in a way that the inflection and derivational layer operate without interfering with each other. Moreover, the analyzer would be required to recognize and then segment the affixes correctly in order to arrive at the root form. This process can be carried out with some accuracy with the use of paradigms. The following section would elaborate on how this is done for the derivational as well as the inflectional suffixes.

3 Paradigm creation using Lttoolbox

The word-and-paradigm approach to inflection lends itself well to analysis within a morphological analysis tool like Lttoolbox. The tool requires the creation of a morphological dictionary that shows correspondences between **surface forms** (SF) and **lexical forms** (LF) (Garrido-Alenda & Forcada 73). SFs are the inflected forms of words that would be found in texts whereas LFs refer to the base forms of those words. For instance, the word ga va la (village + DAT) is the SF of the LF ga va (village). This information is subsequently used to build morphological analysers and generators. The analyzer will take a SF as input to return the LF.

The regularities seen in the correspondences between SF and LF are easily encoded in the form of paradigms. The paradigms are actually rules that are organized in 'blocks' (Stump in Spencer and

Zwicky eds. 38). For inflectional paradigms, the work has already been carried out for Marathi by Akshar Bharati et al. They have created as many as 114 paradigms in order to handle inflectional morphology. However, the existing morph analyzer does not handle all cases of derivational morphology. Moreover, most of these are handled by entering them in the dictionary. By using the Lttoolbox morphological dictionary format, these forms can be analyzed using nested paradigms.

A normal paradigm is created in the morphological dictionary file using a simple XML format. The dictionary file consists of three main parts- symbol definitions (sdefs) where the grammatical features like category, gender and number are defined. This is followed by the paradigm definitions (pardefs) and finally by the lexical entries in the dictionary, which consist of the LF and the paradigm to which they belong. A sample dictionary entry is as follows: (The examples from the dictionary will follow the WX notation for Devanagari. A gloss is provided)

<e lm="kacarA"><i>kacar</i><par n="rasw/A__n"/></e>

kacarA = k tsra (rubbish)
raswA = r sta (road)
Here, the dictionary entry (e lm is entry lemma) is followed by its 'identity' form and the paradigm to which it belongs. The identity form is somewhat identical to the stem of the word. Affixes would be added to this identity form. The paradigm for raswA would be the following:

<pardef n = "rasw/A__n">
   <e>
     <p>
     <l>A</l>
     <r>A</l><s n = "n"/><s n = "sg"/><s n = "parsarg:0"/></r>
     </p>
   </e>
    <e>
    <p>
    <l>yAlA</l>
    <r>A</l><s n = "n"/><s n = "sg"/><s n= "parsarg:lA"/></r>
     </p>
   </e>
</pardef>

Here, the <e> and <p> tags stand for entry and pair respectively. The <l> and <r> tags stand for left and right, which show how the SF and LF should be created from the identity form respectively. In this way, paradigms can be created for all the inflections that a word takes. The existing morph analyzer for Marathi takes into consideration nearly 964 suffixes and postpositions which attach themselves to nouns. This leads to a very large number of SFs being added to each paradigm. This is necessary, however, to recognize the large number of inflectional forms in the language.

In order to deal with a derivational suffix, we use a nested paradigm. For instance, the nested paradigm for the adjective 'l han ' (small) which takes the derivational suffix p a will be as follows:

<pardef n = "lahAna/__a">
   <e>
     <p>

```
       <l></l>
       <r></l><s n = "adj"/></r>
        </p>
     </e>
      <e>
      <p>
      </l>
      </r>
       </p>
       <par n= "D__paNA">
     </e>
  </pardef>
```

The paradigm within lahAna calls the paradigm for paNA and hence is able to recognize a form like 'l han p   a' (smallness).

On the other hand, the paradigm for a noun like neta (leader) will be as follows

```
  <pardef n = "new/A__a">
      <e>
        <p>
        <l>A</l>
        <r>A</l><s n = "nm"/> ><s n = "number:eka"/> <s n = "parsarg:0"/></r>
         </p>
      </e>
  <e>
  <p>
        <l>yAne</l>
        <r>A</l><s n = "nm"/> ><s n = "number:eka"/> <s n = "parsarg:ne"/></r>
         </p>
      </e>
  <e>
   <p>
        <l>yAlA</l>
        <r>A</l><s n = "nm"/> ><s n = "number:eka"/> <s n = "parsarg:lA"/></r>
         </p>
      </e>
        <e>
        <p>
        <l>q</l>
        <r>q</r>
         </p>
         <par n= "D__/wva">
      </e>
  </pardef>
```

The <par n= "D__/wva"> calls the following paradigm:

```
<pardef n = "D__/wva">
    <e>
      <p>
      <l>wva</l>
      <r>A</l><s n = "nm"/> ><s n = "number:eka"/> <s n = "rcat:n"/><s n = "suff:wva"/>
"/><s n= "parsarg: 0"/></</r>
       </p>
    </e>
  <e>
  <p>
      <l>wvAne</l>
      <r>wva</l><s  n  =  "nm"/>  ><s  n  =  "number:eka"/><s  n  =  "rcat:n"/><s  n  =
"suff:wva"/><s n= "parsarg: ne"/></r>
       </p>
    </e>
  </pardef>
```

In the second paradigm, newqwva or 'netrut   ' (leadership) is recognized as well as the inflection that can further be attached to the derived form. Moreover, if the form is productive, it is possible to attach another derivational suffix in the paradigm for the suffix, for instance netrut   + an (capable of leadership). It is also possible to add more than one 'function call' to the original word's paradigm. For instance, the adjective l han   discussed above can take the suffix paNA as well as paNa, and could call two paradigms such as:

```
<pardef n = "lahAna/__a">
    <e>
      <p>
      <l></l>
      <r></l><s n = "adj"/></r>
       </p>
    </e>
     <e>
     <p>
     </l>
     </r>
      </p>
     <par n= "D__paNA">
    </e>
     <e>
     <p>
     </l>
     </r>
      </p>
     <par n= "D__paNa">
    </e>

  </pardef>
```

Here, the form newqwva or 'netrut ' (leadership) can be derived from a single entry for neta in the dictionary. However, in case of nouns being derived from nouns, multiple paradigms need to be made for each masculine, feminine or neuter nouns. This can lead to an explosion in the number of paradigms that need to be created. While integrating both inflectional and derivational dictionaries for analysis, it is probably more useful to have the derivational layer operating alongside the existing dictionary for inflectional forms. Essentially, the derivational dictionary is simply listing productive derivational suffixes and stating the morphophonemic rules that would take care of the mapping between surface forms and lexical forms. However, if needed, the single entry in the dictionary can potentially take care of both the inflectional as well as the derivational forms.

3.1 Derivational paradigms for Marathi

The morphological dictionary prepared for Marathi can potentially generate 450,000 forms. This is quite common for a highly inflectional language and it may be noted that not all these forms occur frequently in the corpus. The total number of dictionary entries is 477 in number. The entries were chosen using test data from the CIIL Marathi corpus. There were approximately 80,000 unique words in a section of this corpus and words containing derivational suffixes were extracted out of the data. These were then compiled and made into dictionary entries in the XML format for Lttoolbox.

Paradigm creation involved the survey of the patterns in the data for rules to capture commonly occurring morphophonemic rules. 38 paradigms were created in order to take care of the commonly occurring morphophonemic changes. There are 10 suffixes being handled at present:

-p a -paNA	Forms abstract nouns from adjectives/nouns Example: god + p a = god p a (sweet + suf=sweetness)
-karI -k ri	Forms personal nouns from common nouns Example: ga +k ri = ga k ri (village + suf =villager)
-mAna -man	Derives personal nouns from common nouns budd i + man = budd i man (intelligence + suf = one who has intelligence
-gAra -gar	Persian suffix that derives an 'agentive' form of noun from common noun. Example: gunha + gar = gunhegar (mistake+suf=one who commits the mistake)
-xAra -dar	Forms personal nouns from common nouns Example: dz mi n + dar = dz mi n dar (land+suff= landlord)
-AlYU -a u	Forms personal nouns from abstract nouns Example : dzop + a u= dzopa u (sleep + suff = person who sleeps a lot)
-vAna - an	Derives personal nouns from common nouns Example: b l + an = b l an (strength + suff =one who is strong)
-wva t	Added to Sanskrit adjectives to derive abstract neuter nouns Example: s m + t = s m t (equal + suf =equality)
-NUka - uk	Abstract noun derived from verbs Example: mir + uk = mir uk (to display + suf=procession)
-wA -ta	Added to Sanskrit adjectives to obtain feminine nouns Example: udar +ta= udar +ta (kind + suf =kindness)

Derivational suffixes included in the dictionary

3.2 Preliminary evaluation

The dictionary created was tested against an evaluation set of exclusively derivational forms created from noun-forming affixes that were taken from the corpus. Against a total of 212 forms, the dictionary could recognize 68 forms, giving it 32 % coverage.

4 Problems and issues

At present, the format of the dictionary does not allow for certain kinds of phenomena to be analyzed effectively. Primarily, the dictionary does not handle prefixes. Certain noun-forming suffixes like be-, atm , a-, etc. cannot be analyzed in this tool without further processing. Secondly, the vriddhi change, where vowels are lengthened on the addition of a suffix is difficult to handle with paradigms. For instance, when the suffix -y added to the noun p   itr   (pure) we get the form pa itry , which is a case of vowel lengthening. One of the proposed solutions is to create a meta-paradigm that would be able to insert the appropriately lengthened vowel in the inflected form. However, this is yet to be tested for Marathi. Finally, there are some noun-forming derivations that are so regular that they are more appropriately treated as cases of inflection. For instance, the gerund forming - e suffix or the suffix - ari regularly applies to most verbs to form the corresponding noun forms. Hence, the analysis of these forms need not be covered by a derivational dictionary.

In future work, the aim is to increase the coverage of the morph analyzer by the addition of new suffixes and dictionary entries as well as integrate the derivational morph dictionary with the existing morphological analyzer for Marathi.

Acknowledgements

References

Marathi morph analyzer. Akshar Bharati et al. 18th Feb 2009 < http://ltrc.iiit.net/showfile.php?filename=onlineServices/morph/index.htm>

Apertium wiki. Apertium online wiki page. 18th Feb 2009 <http://wiki.apertium.org/wiki/Main_Page>

Damle, M.K. *Sastriya Marathi Vyakaran.* Pune: Damodar Sawalram ani company, 1911

Dressler, W.U, W Mayerthaler, O Panagl, and W.U Wurzel. *Leitmotifs in Natural Morphology.* Amsterdam:J Benjamins, 1987

Garrido-Alenda, Alicia, and Forcada Mikel L. 2002. *Comparing nondeterministic and quasideterministic finite-state transducers built from morphological dictionaries*. Procesamiento del lenguaje natural. No. 29 (2002): 73-80

Pandharipande, R. 1997. *Marathi: Descriptive Grammar*. London: Routledge

Sharma, Utpal, Jugal K Kalita and Rajib K Das. 2008. *Acquisition of Morphology of an Indic language from Text corpus*. ACM Transactions on Asian Language Information Processing. Vol 7 Issue 3 (2008): Article no. 9

Spencer Andrew and Arnold Zwicky. *A Handbook of Morphology.* Oxford: Blackwell Publishing, 2001.

Communication through Secret language: A Case Study based on Parayas' Secret Language

Dileep.V
Department of Linguistics, University of Kerala, Kerala, INDIA
dileep_knr@yahoo.co.in

Abstract

Language is the medium or tool for communication but Secret Language also comes under this definition. A secret language is used by a special group to preserve its identity and exclude the outsiders from the communication process. The term refers rather to the social function of a speech form than to any property of its structure.

Parayas is one among the community of scheduled cast in Kerala. In the earlier times Parayas were slaves. Now Parayas are also motivated to attain equality with other casts. They are using secret language for their own special purposes. Parayas keep this language as a variety of their own and never allow others to learn or to use. They hardly reveal the structure of the code nor give the details about the vocabulary or the secret behind the use and development of this variety. The claim that this secret code has a long history and they are using this form for a long time is to be scruinised. The basic structure of the language is that of Malayalam. The phonology, morphology and the syntactic structure of the language variety are almost the same. The most important feature of the language is its vocabulary and it has no script. These have only oral form. The name of the language is different in different region and it has some slight variation. But that are intelligible to them. This language is in the way of death. The youngsters in Paraya community are not initiated to study or to use this language. The aim of the paper is to investigate the dialect variation of this language and compare it with the Malayalam language.

The people who are residing in Kollam (southern district in Kerala) are using the term 'kumita' to refer to a house. But the people who are in Kasaragod (Northernmost district in Kerala) uses the term 'koonta' but this dialect variation does not affect their understanding become it is intelligible to all of them.

The methodology used for this study is field study method for the progress of the work direct contact with the speakers will be done.

1 Introduction

It is believed that Paraya community has existed since the Dravidian age. For which 'Samgham' literature *Puranaanoor* gives strong and valuable evidence about Parayas. Paraya are distributed in Kerala and Tamilnadu Karnataka and Puthuchery. The name Parya is delivered from the Tamil word 'parai', parai means a drum. As certain Parayas act as drummers on marriage, village festivals, government commercial announcements are proclaimed in olden times.Parayas are also referred by the name Sambava. The name Sambava is derived from the word 'samba' samba means Siva. They find the name Sambava more honorific and acceptable.

Paraya is one of the communities among the scheduled caste in Kerala. In the earlier days parayas were slaves and occupied a low status in the society. They were prohibited from entering the house of the upper caste people. The concept of untouchability has manifested itself in so many ways. Now parayas are also motivated to attain equality with other castes. Present generation has improved from the old. They have achieved a greater participation in education, government jobs and other areas. They are not treated as untouchables in any of their relation with the higher castes or at public place, like hotel, temples, churches, etc. Not only that, they were characteristically different from other scheduled castes in Kerala, for instance they were more intelligent and showed great dexterity in martial arts. Traditionally they were engaged in producing baskets from bamboo materials. Even now they are following that traditional job .Nevertheless parayas are also engaged

in agricultural labor, sweeping, etc. Their religious faith was on the basis of ancestral belief and customs. Now they have been completely absorbed in to the Hindu system. Earlier, many Parayas had converted to Christianity but now a days they have begun reconverting to the Hindu fold.

2 Secret Language

Language is the medium or tool for communication but secret language also comes under this definition. A secret language is used by a special group to preserve its identity and to exclude outsiders. Thus the term refers rather to the social function of a speech than to any property of its structure. Occasionally as the Toda language in India, a quite distinct language served the function of secrecy. Most so-called secret language is grammatically the same as the language of the surrounding community, but consists in systematic substitutions of individual words. In this case it is difficult to distinguish clearly between secret language in a strict sense and argot, slang which used ephemeral substitutions and cast, class and technical occupational dialects.

3 Parayas' secret Language

The language variety is used among the Parayas for their special purpose. Paraya keep this language variety as their own and never allow others to learn or use it. Their communities are not willing to teach this language to the other communities. They keep it in their own families. They were using this language for their self protection in olden times. They kept away this language from the mainstream. "*Paraya dialect has been commonly used in relation to the paddy cultivation. It was also used as code language to escape themselves from the attacks of dominant classes*". It is little known to public, other than person who know that they have such a dialect still practiced among them. It kept using secretly. The new generation is completely ignorant about when and where this language originated. There is a mythical belief that this language is used from the time of 'Paakkanaar'. The name of this language is different in different region and it has some slight variation. But that are intelligible. In some places this language is known as '*vaaplanc*' in some other place is referred as '*paRappeecc*'. One important feature of this language is that its unique vocabulary. The code has a special vocabulary of its own which is very much different from Malayalam. Kinship terms are extremely different from Malayalam. The names of flora, fauna, and house hold items and body parts also posses a different type of vocabulary structure. The most important feature of this language is that it has no script.

4 Aim and objective of the study

1. To examine the structure of Parayas secret language.
2. To analyze the surroundings which lead to the development and usage of the secret language?
3. To bring out the dialect variation of the language
4. Compare all the above with Malayalam

5 Methodology

The methodology used for this study was field study method. For the progress of the work direct contact with the speakers was done. People belonging to different age groups and who belonged to the educated and non educated class were selected for this study. The area which I have selected for this study is the Kollam and Thiruvananthapuram district in Kerala.

6 Analysis of the data

217

Dialect variation

Dialect variation is common to all the languages in the world. Each language varies according to regional differentiation. For instance two different regions use separate word to represent the same meaning, but they would understand the differences in the usage. Likewise this Paraya script also has regional variation in the word usage.

Eg		Malayalam
kumita	house	viit
koonta	house	viit

The people who reside in Kollam district use the term *'kumita'* to refer to house but the people who are in Thiruvananthapuram district use the term *'koonta'*. Similarly there are many words in this language

navaa	speak	paRayuka
meRRappa tt	speak	paRayuka
anump	alcohol	madhyam
tu ump	alcohol	madhyam
taamaR	police	police
kuuRump	police	police
kelali	face	mukham
mohamaaR	face	mukham

Comparative analysis with Malayalam

This language use several Malayalam words, but it differ in their meaning when applied in the secret language.

    'mayakkuka'   to come

In this secret language the word 'mayakkuka' is applied to denote the word 'to come' but in Malayalam it is used to make sedate somebody.

    'kooccuka'   throw

In Malayalam 'kooccuka' means to feel cold because of winter season.

    'ceRayuka'   died

 In Malayalam it means to because angry with body bending to side

    'minni'   upees

In Malayalam it denotes sparkling, lightening.

Markers

The secret dialect uses the tense, case, gender markers that are similar to that of Malayalam.

Tense

/unnu /, /u, i /, um/ these are the tense markers in Malayalam

/unnu/ - is the present tense markers

    meekk unnu   to drink
    cerRakk unnu   to kill

/ i, u/ - is the past tense Marker

    meek i   drank

ceRacc u   killed

/um / -is the future tense marker
        meek um   will drink
        ceRakk um   will kill

Case
/e/, /oo t/, /u /, / il / these are the case markers in Malayalam
         'pikkin e'   him (accusative)

'pikk' means a person or he, 'pikkine' means him. The case marker 'e' is similar to that of the accusative case marker in Malayalam.
        'pikkin oo '   with him (sociative)
         oot - is the sociative case marker

        'pikkan u'   to him (dative)
        u - is the dative case markers

        'pikkan il'   in him (locative)
        il - is the locative case marker

Gender
/an/, /atti /- are the masculine, feminine gender markers in Malayalam
        Paa uv an   Paraya man
        Kalins an   Pulaya man

        Paluv atti   Paraya women
        Kalinn atti   Pulaya women

Numerals
Another important resemblance is in the field of numerals. The number system both ordinals and cardinals are almost same as the Malayalam numerals.
        onnu   one
        rantu   two

        onnaamatte   first
        rantaamatte   second

Sentence Structure
The basic sentence structure of this language is same that of Malayalam, it follows the SOV order
        raaman raavanane ceraccu (raaman raavanane konnu)
        (Rama   Ravana   killed)
        Rama killed Ravana

        raaman anump meekkum (raaman kallu kudikkum)
        (Rama   alcohol will drink)
        Rama will drink alcohol

        minnippaat ilunna (kaasu illa)
        (Money   no)
        No money

7 Present situation of this language

This language is in the way of death. Their communities are not willing to teach this language to other communities or person. The narrow minded men of them lead this language to death. Another reason for this condition is the youngsters in Paraya community are not initiated to study or use this language. So it is vanishing from the adversity at parayas. Perhaps educated people among them hesitate to use this dialect in public because it brings to them social inferiority in front of others.

8 Conclusion

This paper tries to analyze the secret language used by the Paraya community. They hardly reveal the structure of the code nor give the details about the vocabulary or the secret behind the use and development of this variety. There is claim that the secret code has a long history. They are using this form for a long time. The basic structure of the language is Malayalam. The phonology, morphology, and the syntactic structure of the language variety are almost the same as that of the language of the state. As this study could collect data from only fifteen informants, it could not claim that this is the detailed study of the variety. A detailed analysis collecting more data and their socio-linguistic features may give a flamboyant picture of the verity. Of course this study stresses that clear documentation of the code is a must for the knowledge world.

Author Details

Attar, Mahan
Mahan Attar is PhD student at the department of English, University of Pune. She has received her M.A. and B.A. degrees in English Language Teaching from Iran. She has taught English for 10 years at pre-universities in Iran and now, her guide is Dr. S. S. Chopra.

Baraik, Sunil
Sunil Baraik is a Ph.D. scholar at the Department of Linguistics, Ranchi University, Ranchi.

Batra, Ridhima
At the time of the paper presentation Ridhima Batra was an MSc student at All India Institute of Speech and Hearing (AIISH), Mysore.

Chengupta, Shyamala
Dr. Shyamala. K. C is currently Professor in Language Pathology at Department of Speech – Language Pathology and currently she is HOD of Department of Material development at All India Institute of Speech and Hearing, Mysore, Karnataka, India. She has more than 20 years of experience in the field of Speech – Language Pathology and has done more than 100 research papers which are published/ presented at national/ international conferences and journals. She has guided 5 doctoral theses successfully and around 50 master's dissertations at All India Institute of Speech and Hearing. She has published 6 books and has written several chapters in international text books. Her area of interest is child and adult language disorders and bi/multilingualism. She has delivered several guest lectures in USA, Germany. She has done around 10 projects at All India Institute of Speech and Hearing in collaboration with Central Institute of Indian Languages, Mysore and Indian Council of Medical research, etc.

Deepa, M.S.
Deepa M.S. is a lecturer at JSS D JSS Institute of Speech and Hearing, Mysore.

Deepthi M
At the time of the paper presentation Deepthi M. was an MSc student at the Department of Speech and Language Pathology, JSS Institute of Speech and Hearing, Mysore.

Devika, M.R.
At the time of presenting the paper, Deivka was an MSc student in Speech Pathology at All India Institute of Speech and Hearing, Mysore.

Dharurkar, Chinmay Vijay
Chinmay is a PhD scholar at Indian Institute of Technology, Bombay, Mumbai.

Gopal, Kusum
Kusum Gopal has M.A. in Hindi from SCD Govt. College, Ludhiana.

Harshan Kumar H.S.
Harshan is a BSc student at JSS Institute of Speech and Hearing.

Jadhav, Arvind Tukaram
Arvind Jadhav is a lecturer in English at Yashwantrao Chavan College of Science, Karad, Kolhapur Maharashtra.

Kak, Aadil Amin
Aadil Amin Kak is associate professor at the Department of Linguistics, University of Kashmir, Srinagar.

Karmakar, Samir
Samir Karmakar is a PhD scholar at the Indian Instute of Technology, Kanpur.

Kumar, Ritesh
Ritesh Kumar is a PhD scholar at Centre for Linguistics, Jawaharlal Nehru University, New Delhi.

Kumar, Santosh
Santosh Kumar is a Ph.D. scholar at the Department of Linguistics, University of Delhi, Delhi.

Kumar, Shailendra
Shailendra Kumar is a PhD scholar at the Department of Linguistics, Banaras Hindu University. His research interests are Hindi-English divergence issues and particles. He has done MA in Linguistics from Banaras Hindu University.

Lahiri, Bornini
Bornini Lahiri is doing M.Phil in linguistics at J.N.U, New Delhi. She completed her masters in linguistics in the year 2008 from J.N.U. Her area of interest is typology and areal linguistics.

Lawaye, Aadil Ahmed
Aadil Ahmad Lawaye is a research scholar in linguistics at the Department of Linguistics, University of Kashmir, Srinagar.

Lone, Feroz Ahmad
Feroz Ahmad Loneis a research scholar in linguistics at the Department of Linguistics, University of Kashmir, Srinagar.

Mala, Christopher
Christopher is a PhD scholar at International Institute of Information Technology, Hyderabad.

Malik, Pallavi
At the time of the paper presentation Pallavi Malik was an MSc student at All India Institute of Speech and Hearing (AIISH), Mysore.

Mayee, P. Kiran
Kiran Mayee is a Ph.D. scholar at International Institue of Information Technology, Hyderabad.

Mehdi, Nazima
Nazima Mehdri is a research scholar in linguistics at the Department of Linguistics, University of Kashmir, Srinagar.

Mukherjee, Aparna
Aparna Mukherjee is a research scholar in Linguistics at the Centre for Linguistics, Jawaharlal Nehru University, New Delhi.

Naidu, Vishwanath
Vishwanath Naidu is a PhD scholar at International Institue of Information Technology, Hyderabad.

Nanda, Dubi Dhakal
Dubi Dhakal Nanda is a PhD scholar at Tribhuvan University, Kathmando, Nepal.

Nash, Joshua
Joshua Nash is a PhD scholar at University of Adelaide, Austrailia.

Navitha U.
At the time of presenting the paper, Navitha was an MSc student in Speech Pathology at All India Institute of Speech and Hearing, Mysore.

Nayak, Madhu K.
Madhu is an internship student at JSS Institute of Speech and Hearing, Mysore.

Ozarkar, Renuka
Renuka Ozarkar is a research scholar at the Department of Linguistics, University of Mumbai, Mumbai.

Ravi, Sunil Kumar
Mr. Sunil Kumar. R is currently working as Research Officer at Department of Speech Language Pathology at All India Institute of Speech and Hearing, Mysore, India. His area of interest includes adult language disorders, child language disorders and has done around 10 research papers in this area. He has published several articles in national and international journals like Asia Pacific Journal of Speech, Language, Hearing (APJSLH); International Journal of Dravidian Linguistics (IJDL); Interdisciplinary Journal of Linguistics (IJL); Language in India (Online Journal), Journal of Indian Speech and Hearing Association (JISHA). He has also presented several research papers in national and international conferences like Annual Conference of Indian Speech and Hearing Association (ISHA Conference), All India Conference on Dravidian Linguistics (AICDL); International Conference on South Asian Languages (ICOSAL), etc. He has good academic, clinical and research skills in the area of Speech – Language Pathology.

Saha, Atanu
Atanu Saha is a PhD scholar at the Centre for Linguistics, SLL&CS, Jawaharlal Nehru Unversity, New Delhi.

Saurov, Syed
Syed Saurov is a research scholar at the School of Language Sciences, English and Foreign Language University, Hyderabad.

Shekhar, Sudhanshu
Sudhanshu Shekhar is a PhD scholar at the Centre for Linguistics, SLL&CS, Jawaharlal Nehru Unversity, New Delhi.

Shwetha MP
At the time of the paper presentation Shwetha MP was an MSc student at the Department of Speech and Language Pathology, JSS Institute of Speech and Hearing, Mysore.

Suhas K.
Suhas is a BSc student at JSS Institute of Speech and Hearing.

Swapna, N
Swapna is at All India Institute of Speech and Hearing, Mysore.

Thakur, Gayetri
Gayetri Thakur has done M.A. in Linguistics from Banaras Hindu University, Varanasi and is pursuing research in linguistics.

Trupthi T
At the time of the paper presentation Trupthi T. was an MSc student at the Department of Speech and Language Pathology, JSS Institute of Speech and Hearing, Mysore.

Vaidya, Ashwini
Ashwini Vaidya's academic background with respect to Linguistics started during her Masters at Mumbai University in 2003. After completing her degree with distinction, she worked for a year in a technology start-up. Impressed by the possibilities of applying linguistics for the creation of powerful tools for language analysis, she decided to pursue her M. Phil. in Computational Linguistics. At the time of presenting the paper, she was at IIIT-Hyderabad in her final semester.

Vamanan, Dileep
Dileep is a research scholar at Department of Linguistics, University of Kerala, Thrivendram.

Vashistha, Neha
Neha Vashistha is a PhD scholar at the Department of Linguistics, Banaras Hindu University. Her research interests are particle disambiguation and contrastive grammars of Hindi-English. She has done MA in Linguistics from Banaras Hindu University.

Team SCONLI-3
Abstract Review Committee

Faculty Members
Prof. Anvita Abbi
Prof. Vaishna Narang
Prof. Proamod K.S. Pandey
Dr. Ayesha Kidwai
Dr. Girish Nath Jha

Student Members
Narayan Choudhary
Gibu Sabu M.
Maansi Sharma
Pooja M.

Organizing Committee

Faculty Coordinator: Dr. Ayesha Kidwai
Student Coordinator: Narayan Choudhary
Student Coordinator: Gibu Sabu M.
Student Coordinator: Maansi Sharma

Members: Smita Joseph
            Meiraba Takhelambam
            Gopal Parihar
            Atanu Saha
            Ritesh Kumar
            Sudhanshu Shekhar

Committees

Registration
- Pooja (Head) – M.Phil.
- Maansi Sharma– Ph.D.
- Rajlakshmi – M.A.
- Rajtilak– M.A.
- Madhav Gopal – M.A.

Food
- Gibu Sabu M. (Head) – Ph. D
- Meiraba Takhelambam– M.Phil.
- Atanu Saha – M.Phil.
- Bornini Lahiri – M.Phil.
- Ranjini Mazumdar– M.Phil.
- Sweta Sinha – Ph.D.

Xerox
- Ritesh Kumar (Head) – M.Phil.
- Atanu Saha – M.Phil.
- Mahesh– Ph.D.
- Prakash– M.A.
- Vivek– M.A.
- Sujoy Sarkar – M.A.

Publicity
- Narayan Choudhary (Head) – Ph.D.
- Smita Joseph – Ph.D.
- Devina Kaul– M.Phil.
- Meiraba Takhelambam– M.Phil.
- Sudhanshu Shekhar– M.Phil.

Accommodation
- Meiraba (Head - Boys) – M.Phil.
- Neha Mishra (Head - Girls) – M.Phil
- Narayan Choudhary – Ph.D.
- Maansi Sharma– Ph.D.
- Pooja M. – M.Phil.
- Radhika Gopalkrishnan – Ph.D.
- Preeti Rekha Gogoi – Ph.D.
- Atanu Saha – M.Phil.
- Mahesh M. -Ph.D.
- Ritesh Kumar – M.Phil.
- Akash Raha -M.A.
- Dev Bharati – M.A.

Hospitality
- Ranjini Mazumdar – M.Phil.
- Sabiha Hashami – M.Phil.
- Moumita Dey – M.Phil.
- Reshmi P– M.Phil.

Session Chairing
- Sabiha Hashami – M.Phil.
- Prerna Kathuria – M.Phil.
- Kulsum Mehwish – M.A.
- Deepa Thomas – M.A.
- Janini Kandhadai – M.A.
- Hima S. – M.A.
- Nimmi – M.A.
- Sujoy Sarkar – M.A.