Nepalese Linguistics

Editor-in-Chief   Kamal Poudel

Editors   Ram Raj Lohani
                               Dr. Tikaram Poudel

Editorial

Linguistic Society of Nepal, since its inception in 1979, has been involved in preserving and promoting the languages of the Himalayan region through different activities such as organizing conferences, workshops and publications. As all our esteemed readers know that the journal *Nepalese Linguistics* is one of the major initiatives of the Society. The Board of Editors feels immense pleasure to bring out Volume 33.1 of *Nepalese Linguistics* in the eve of the 39th International Annual Conference of Linguistic Society of Nepal.

The Society decided to peer-review the articles since this issue in order to ensure the quality of the journal. We created a pool of reviewers of more than thirty scholars from the various areas of linguistics. We are grateful to the contributions of the reviewers, who accepted our requests and reviewed the articles within short period of time. We sincerely acknowledge their support. Similarly, we thank the authors for submitting their articlesand revising them to address the comments made by the reviewers. The consent of the writers and reviewers has encouraged us to pave a new beginning of the journal.

We are aware that the time allocated to the writers and reviewers was not sufficient, and we assure you all that the editorial board will embark on its activitiesin time in the future issues.In spite of our sincere effort, we could not accommodate all the articles submitted to the editorial board. The articles that the authors revised following the reviewers' suggestions reached us just in the eve, and some are yet to be received. Similarly, we are still waiting for feedback on some articles.Thanks to the limitation of time; these articles could not obviously be included in this issue. The board will immediately proceed ahead for the publication ofthe second issue of this volume.

The current issue comprises eight articles and the keynote address delivered at the 38th Conference of the Society. These articles cover five different themes namely morphosyntax, sociolinguistics, computational linguistics, acoustic phonetics and language planning.

We acknowledge the cooperation of the executive members of Linguistic Society of Nepal, authors, reviewers and other individuals and organizations involved in the publication of the journal. The Board specially recognises the support extended by Mr. Krishna Prasad Chalise in preparation of this issue. Finally, we look forward to constructive feedbacks from our esteemed readers.

Board of Editors

CONTENTS

Pitambar Behera is affiliated to Jawaharlal Nehru University. The author's email address is <pitambarbehera2@gmail.com>.

Sharmin Muzaffar is affiliated to Aligarh Muslim University. The author's email address is <sharmin.muzaffar@gmail.com>.

PushpaRenu Bhattacharyya is affiliated to Tezpur University, India. The author's email address is <dr.pushparb@gmail.com>.

Krishna Prasad Chalise is affiliated to Tribhuvan University, Nepal. The author's email address is <krishnaprasadchalise@gmail.com>.

Kuntala Ghosh Dastidar is affiliated to University of Calcutta, India. The author's email address is <ghoshdastidar.kuntala28@gmail.com>.

Cathryn Donohue is affiliated to Hongkong University, Hongkong. The author's email address is < cdonohue@gmail.com>.

Bhim Lal Gautam is affiliated to Central Department of Linguistics, Tribhuvan University, Nepal. The author's email address is <gautambhim@rocketmail.com>.

Dan Raj Regmi is affiliated to Central Department of Linguistics, Tribhuvan University, Nepal. The author's email address is <danrajregmi8@gmail.com>.

Ruth Rymbai is affiliated to North Eastern Hill University, Shillong, India. The author's email address is <rymbairuth231986@gmail.com>.

Arvind Kumar Rawat is affiliated to North Eastern Hill University. The author's email address is <arvind.rawats12@gmail.com>.

Tanmoy Bhattacharya is affiliated to University of Delhi, India. The author's email address is <tanmoy1@gmail.com>.

DEVELOPING CLASSIFICATION-BASED NAMED ENTITY RECOGNIZERS (NER) FOR SAMBALPURI AND ODIA APPLYING SUPPORT VECTOR MACHINES (SVM)

Pitambar Behera and Sharmin Muzaffar

This paper demonstrates the development of named Entity Recognizers (NER) applying Support Vector Machines (SVM) for Sambalpuri and Odia. The Sambalpuri corpus amounts to 112k word tokens out of which 5,887 are named entities. On the contrary, 250k ILCI corpus has been applied for Odia out of which 18,447 tokens are named entities. The former accurately recognizes 96.72% whereas the latter provides 98.10% accuracy.

Keywords: *NER, Sambalpuri, NLP, Odia, SVM, Machine Learning, Indo-Aryan languages, Information Retrieval, Natural Language Processing.*

1 Overview

Named entity recognition (NER) is one of the applications of Natural Language Processing and it is considered as the subtask of information retrieval. NER is the process of detecting Named Entities (NEs) in a document and to categorize them into certain named entity classes such as the names of organization, person, location, sport, river, city, country, quantity etc. In English, we have accomplished a lot of work pertaining to recognizing named entities. On the contrary, we have not achieved remarkable accomplishment with regard to detecting NER in Indian languages. India is an abode for 22 official languages along with endangered, lesser-known and less-studied languages. NER is still considered to be an emerging area of research in the field of NLP in the context of Indian languages.

There are various applications of NER such as Information Extraction, Question Answering, Information Retrieval, Automatic Summarization, Machine Translation, etc. The Named Entities can be made known to us by performing computation on a given natural language through rule-based or statistical approaches. The task of identification, extraction and retrieving necessary information can be made faster, if we are already acquainted with the nature, type and functions of named entities. Therefore, NER is the process of detecting, classifying and extracting Named Entities in a document into their corresponding classes with the application of any of the NER based approaches.

1.1 Approaches to named entity recognition

There are basically two broad approaches that are employed in the recognition of named entities (Nayan et al., 2008; Sasidhar et al., 2011; Saha, 2008). These include: Rule-based approach and Machine learning based approach (Kaur and Gupta, 2012; Kaur and Gupta, 2010; Srivastava et al., 2011).

1.1.1 Rule-based approach

Under this section, there are list lookup approach and linguistic approach. So far as the former is concerned, gazetteers are exploited that comprise of different lists of named entity classes and a simple look up or search operation is conducted in order to detect whether a word belongs to a named entity class or not. If a particular word belongs to a named entity class, a named entity label, as specified in the annotation schema, is allotted to that word on the basis of the named entity class which it originally belongs to. On the other hand, in linguistic approach, a linguist is entrusted with the work of formulating heuristic linguistic rules, so that the named entities can be identified as well as classified and extracted easily (Ekbal and Bandyopadyay, 2010; Gupta and Lehal, 2011). The formulated rules are language dependent and cannot be applied in order to identify named entities in any other given language (Kaur and Gupta, 2012). Therefore, data-driven statistical approach became indispensable.

1.1.2 Statistical approach

This approach is motivated by the machine learning theories and algorithms, for instance, Hidden Markov Models (HMM), Maximum Entropy Markov Model (MEMM), Conditional Random Field (CRF), Support Vector Machines (SVMs), Decision Tree and so on.

2 Review of literature

For English, we have developed rule-based NER systems of which the F-measure accuracy ranges from 88-92% (Grishman, 1995; Wakao et al., 1996). These rule-based systems are inoperable and they demand huge amount of linguistic knowledge of the given language. Machine Learning based techniques are operable in nature and does not require huge linguistic knowledge in that language.

Hidden Markov (Bikel et al., 1997), Maximum Entropy (Borthwick), CRF (Li and Mccallum, 2004) models have been commonly applied to different languages for NER identification purposes. A hybrid system combining ME, HMM and rule-based methods has been developed by Srihari et al. (2000).

Gali et al (2008) have reported lexical F-Score accuracy of 40.63%, 39.04%, 40.94%, 43.46% and 50.06% for Bengali, Oriya, Telugu, Urdu and Hindi respectively. For Indian languages, Ekbal and Bandopadhyay (2007) have described an approach to lexical pattern learning. For Hindi and four other European languages, Cucerzan and Yarowsky (1999) have developed NER applying morphological and contextual features. They registered an f-measure of 41.70%. Li and Mccallum (2004) have developed an NER for Hindi applying CRFs and registered an f-measure of 71.50. Kumar and Bhattacharya (2006) have achieved an accuracy of 79.7% f-measure on a maximum entropy markov model for Hindi. Krishnarao et al (2007) have demonstrated a comparative analysis of CRF and SVM for the purpose of recognizing named entities in Hindi. Chopra et al (2012) have prepared an NER for Hindi combining rule-based heuristics and HMM where they have achieved 94.61% accuracy.

Shishtala et al. (2003) have developed an NER for Telugu applying CRF and registered a reported F-value of 44.91%. Ekbal and Bandyopadhyay (2008) have developed an NER for Bengali applying SVM and reported an F-Score of 91.8%. Kaur and others have developed an NER for Punjabi language applying CRF. Balabantaray and others (2013) have developed an NER in Odia having parameterized CRF++ tool in various ways. In doing so, they have applied the gazetteer and parts of speech tags so as to extract various feature sets. So far as the corpus distribution is concerned, they have applied 45k of word tokens data out of which only 1k are named entities. They have achieved an overall precision value of 92%.

3 Support vector machines

Support vector machines are a group of supervised learning models developed by Vapnik (Joachims, 1999). They are associated with learning algorithms that are responsible for analyzing data and recognizing patterns that are further applied to the purpose of classification and regression analysis tasks. When you provide a finite set of training data by marking each as belonging to either of the categories, the SVM training algorithm develops a model. It provides labels to new examples learned from the training examples, making it a non-probabilistic binary linear classifier.

The SVM learns a linear hyperplane when given a particular series of N training examples $\{(x1, y1),…, (x_N, y_N)\}$ where an example x_i stands for a vector R^N and the class annotation label is $y_i \in \{-1,+1\}$. The hyperplane distincly separates positive instances from the set of negative ones with an optimum margin. That maximal margin is defined as the distance of the hyperplane to the nearest of the positive instance and negative set of examples (Gimenenez and Marquez, 2006).

A non-linear classifier decides $f(x)=\text{sign}(g(x))$ for an input vector where $f(x)=+1$ (x is a member of a given class), $f(x)=-1$ (x is not a member of a given class). $g(x)$ is proportionate to m, z_i s are support vectors and
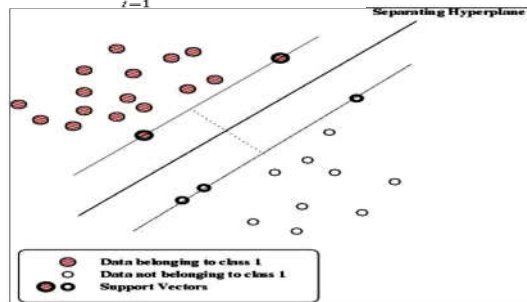
$$g(x) = \sum_{i=1}^{m} w_i K(x, z_i) + b$$

Fig 1: SVM example of hard margin classifying negative and positive examples

K stands for kernel.

Given some training data $\mathcal{D}$, a set of n points of the form

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^p, \ y_i \in \{-1, 1\}\}_{i=1}^{n} \quad (4)$$

4 Methodology

4.1 Corpus annotation schema

Concepts are categorized in three broad categories: Location, Organization, and Person. But we have categorized those into seven general categories so as to incorporate all of them. Entities of person are the names of persons living or dead, of deities, of fictional characters etc. For instance, Hari, Gangadhar Meher, Kalapahada. Organization entities are restricted to institutions, corporations and government agencies such as Sambalpur University, DRDO, etc. Similarly, the entities pertaining to location are countries, streets, mountains, airports, monuments such as Sambalpur, Jharsuguda Airport, Bir Surendra Sai Statue. Temporal entities are the names for months, weeks, special days as for example September, Independence Day. Date Entities are the entities suggesting the dates of a particular month. Entities referring to currency names are Rouble, Yen, Rupee, Yuan etc. Names referring to percentage and fractional numbers are clubbed under this category.

The annotation schema (see fig 2) consists of seven labels such as person (NNP_PER), organization (NNP_ORG), location (NNP_LOC), time (NN_TIM), date (NN_DAT), currency (CD_VAL), and percentage (CD_PER). We have developed our own annotation schema considering various lexical feature sets of named entities in Sambalpuri and Odia.

Fig 2: NER annotation schema

The figure demonstrated below explains the examples of named entities pertaining to the broad category of location.
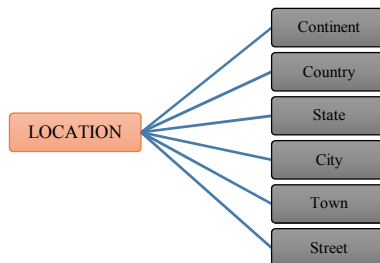
Fig 3: Examples of location named entity

Table 1 below contains the annotation samples of Sambalpuri NER data as according to their respective categories.

Table 1: Sambalpuri annotation sample

Labels	Sambalpuri Annotation Sample
NNP_LOC	bʰɛɽen NNP_LOC
NNP_NAM	səməlpurɪ NNP_NAM
NNP_ORG	səməlpur NNP_ORG ɪnbʰɑ:rsɪʈɪ NNP_ORG
NNP_PER	pəndɪ̈ʈ NNP_PER josi: NNP_PER
NN_TIM	sɛpʈembər NN_TIM
NN_DAT	12.09.2018 NN_DAT
CD_VAL	dɔlɑ:r CD_VAL
CD_PER	12% CD_PER

4.2 Corpus distribution

For the current study, we have taken 112k and 250k data for Sambalpuri and Odia respectively. Out of those aforementioned data, 5887 and 18, 447 have been labelled as NERs in Sambalpuri and Odia respectively. For the purpose of testing, we have applied 2503 word tokens for Sambalpuri and 3003 tokens for Odia.

The corpora for Sambalpuri have been adapted from (Behera et al., 2018); created for POS annotation work. On the other hand, 250k ILCI corpora have been applied for Odia out of which 18,447 tokens are named entities. The POS-annotated corpora for Odia have been adapted from Behera (2016).

Table 2: Corpus distribution

Languages	Training Corpus	Testing Corpus
Sambalpuri	5887/112k	2503
Odia	18, 447/250k	3003

4.3 Feature extraction

The features for a classification-based NER have been selected considering the word tokens, POS, ambiguity and maybe's. The tri-gram feature file has been applied. The configuration file applied in the learning phase encapsulates medium verbose (-V 2) and the directions of automatic learning and annotation have been set to the left-right-left (LRL) mode. All the miscellaneous features have been set to the default mode.

```
# SVMT configuration fileNAME = /home/sanskrit/svmtool/models/odi/ODI
TRAINSET =/home/sanskrit/svmtool/odia.trainSVMDIR
=/home/sanskrit/svmtool/svmlight/W = 5 2 F = 5 10000 X = 7 Dratio = 0.005
REMOVE_FILES = 1do M0 LRL#do M1 LRL#do M2 LRL#do M4 LRL#
....................................................
-------#ambiguous-right [default]A0 = w(-3) w(-2) w(-1) w(0) w(1) w(2) w
(3) w(-2,-1) w(-1,0) w(0,1) w(-1,1) w(1,2) w(-2,-1,0) w(-2,-1,1) w(-1,0,1)
w(-1,1,2) w(0,1,2) p(-3) p(-2) p(-1) p(-2,-1) p(-1,1) p(1,2) p(-2,-1,1) p
(-1,1,2) a(0) a(1) a(2) a(3) m(0) m(1) m(2) m(3) z(2) z(3) z(4) ca(1) cz
(1)A0unk = w(-3) w(-2) w(-1) w(0) w(1) w(2) w(3) w(-2,-1) w(-1,0) w(0,1) w
(-1,1) w(1,2) w(-2,-1,0) w(-2,-1,1) w(-1,0,1) w(-1,1,2) w(0,1,2) p(-3) p(-
2) p(-1) p(-2,-1) p(-1,1) p(1,2) p(-2,-1,1) p(-1,1,2) k(0) k(1) k(2) k(3)
m(0) m(1) m(2) m(3) a(2) a(3) a(4) z(2) z(3) z(4) ca(1) cz(1) L SA AA SN CA
CAA CP CC CN MW#
```

Fig 4: The configurations file for Sambalpuri and Odia NERs

5 Evaluation

The statistical representation in the below tabulated data (see Table 3 and Table 4) demonstrates the evaluation of the NER systems for Sambalpuri and Odia category-wise on three measures: perecntage, precision and recall. Both in Table 3 and Table 4, the highest accuracy is figured in the categories of proper names, person, location, and organization which is indicative of the fact that at the level of POS category of proper nouns, the Sambalpuri (Behera and Dash, 2017) POS tagger (Behera et al., 2018) and Odia (Jha et al., 2014) POS tagger (Behera, 2016, Behera, 2017, Ojha et al., 2015) have also registered a fair amount of accuracy. On all the measures, cent percent accuracy has been achieved in the cattegory of named enitity of time. In both these systems, named entity of persons has registered almost the same amount of accuracy as explained in the following comparative data. The zero figured in Table 3 and Table 4 refers to the fact

that during evaluation the test file does not contain any word for the respective categories.

Table 3: Evaluation of Sambalpuri NER

Labels	(%)	precision	recall
NNP_LOC	92.15	96.79	92.15
NNP_NAM	97.60	100	100
NNP_ORG	91.94	93.71	94.17
NNP_PER	99.09	98.60	99.09
NN_TIM	100	100	100
NN_DAT	0	0	0
CD_VAL	100	100	100
CD_PER	0	0	0
Total	96.72	72.62	73.17

Table 4: Evaluation of Odia NER

Labels	(%)	Precision	Recall
NNP_LOC	99.04	98.57	99.04
NNP_NAM	89.83	100	100
NNP_ORG	97.13	96.73	97.13
NNP_PER	99.03	97.73	99.03
NN_TIM	100	100	100
NN_DAT	100	100	100
CD_VAL	83.33	50	100
CD_PER	0	0	0
Total	98.10	80.37	86.90

6 Computational framework

This section highlights the computational architecture of the user interface (see Fig. 5). At the first stage, raw text in Sambalpuri and Odia is provided. Thereafter, the system pre-processes the junk characters and avoids unwanted elements in the text. At the third stage, the texts are tokenized. Then, the text goes to the SVM tool where it is

POS-tagged after being pre-processed. The POS-tagged text is again annotated with NER annotation labels. The text gets detokenized at the detokenization level. Finally the system provides the NER-tagged output after detokenization.
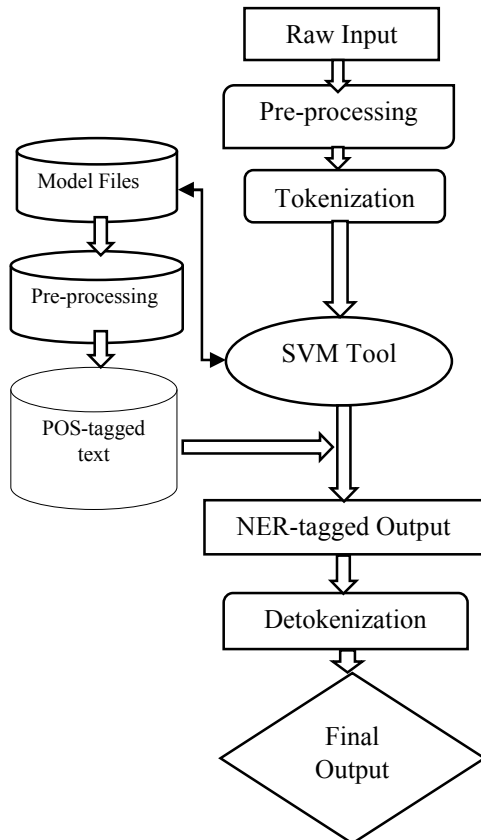

Fig 5: User interface architecture

7   Issues and challenges

One of the major issues faced by Indian languages written in their respective scripts is that there is no capitalization applied for the proper named entities at the initial alphabet unlike English. Therefore, it becomes quite challenging in order to detect the entities properly. The second challenging issue is that they have both inflectional and agglutinating properties, enriched morphological features and relatively free word order. Another inherent linguistic issue to be pondered upon is that they use several common nouns as the proper nouns for example, Pawan, Chandni, Dharti. In addition, dictionaries applied for the recognition task of named entities also contain the above issue that makes the detection process even more complicated. Furthermore, another issue related to the linguistic challenge is the lack of standardization in spelling, especially for the non-scheduled languages such as Sambalpuri, Bhojpuri, Magahi, Awadhi and many others.

The web has been dominated by the English and Mandarin languages. Above all, we need to accept the fact that even after years of endeavor by the CIIL, TDIL, IIIT Hyderabad and other institutions, the resource scarcity situation has not changed drastically so far. As a result, we don't have properly labelled data for NER with standaridization. Owing to the scarcity in resource, we do not have name dictionaries, POS taggers, morphological analyzers, gazetteer lists and so on that can be utilized for the development of NER systems.

So, the need of the hour is to make the Indian languages resource rich by making them available on the digital space. Secondly, we need to develop unambiguous dictionaries clearly distinguishing the proper noun and common noun labels for the same word as exemplified aforementioned. To address this issue, we need to develop more frequency dictionaries which can predict the frequency of the co-occurrence of the proper with the common nouns. The phonetic characteristic of the Brahmi script of the Indian languages can be exploited by the NER systems in order to get the desired output.

8   Conclusion

The rationale for dealing with Sambalpuri and Odia simultaneously is that they are considered to be closely related independent languages. We have presented two classification-based SVM NERs for Odia and Sambalpuri in this study. In addition, we have also applied the POS taggers for respective languages during detection process of named entities. We have achieved 96.72% and 98.10% accuracy in Sambalpuri and Odia respectively. We have evaluated Sambalpuri 72.62/73.17 and Odia 80.37/86.90 in precision and recall methods respectively. Our system performs better than already reported Odia NER. The NER system for Sambalpuri is the first

attempt at successfully recognizing named entities for the language.

This research work is limited to only the identification and extraction of named entities without nested structure. NERs will be fruitful for Information Extraction, Question Answering, Information Retrieval, Automatic Summarization, Machine Translation etc. in Sambalpuri and Odia for further future research.

Acknowledgements

References

Balabantaray, Rakesh Chandra; Das, Suprava; and Mishra, Kshirabdhi Tanya. 2013. Case study of named entity recognition in Odia using CRF++ tool. *IJACSA*, 4 (6), 213-216.

Behera, Pitambar. 2016. Evaluation of SVM-based automatic parts of speech tagger for Odia. In *Proceedings of WILDRE-3 (LREC-2016), ,* Portoroz, Slovenia, 32-38.

Behera, Pitambar. 2017. An experimentation with the CRF++ parts of speech tagger for Odia. *Language in India*. Vol 17:1, Jan, (1930-2940), Bloomington, USA.

Behera, Pitambar; Ojha, Atul Kumar; & Jha, Girish Nath. 2015. Issues and challenges in developing statistical pos taggers for Sambalpuri. Human Language Technology: Challenges for Computer Science and Linguistics, ed. By Zygmut Vetulani et al. 393-406. LNAI 10930, Springer International Publishing.

Behera, Pitambar. & Dash, Biswanandan. 2017. Documenting Sambalpuri-Kosli: The Case of a Less-resourced Language. *Indian Journal of Applied Linguistics (IJOAL).* Bahri Publications, 43(1-2), 128-144.

Chopra, Deepti; Jahan, Nusrat; & Morwal, Sudha. 2012. Hindi named entity recognition by aggregating rule based heuristics and Hidden Markov Model. International Journal of Information, 2(6).

Ekbal, Asif & Bandyopadhyay, Sivaji. 2010. Named entity recognition using Support Vector Machine : A language independent approach. International Journal of Electrical and Electronics Engineering, 4:2.

Gimenez, Jesus &Marquez, Lluis. 2006. SVMTool technical manual v1.3, Barcelona: TALP Research Center, LSI Department, Univeritat Plotecnica de Cataluniya.

Gupta, Vishal & Lehal, Gurpreet Singh. 2011. Named entity recognition for Punjabi language text summarization. International Journal of Computer Applications, 33:3.

Joachims, Thorsten. (1999). Making large scale SVM learning practical. In: Schölkopf B, Burges C, Smola A (eds) Advances in kernel methods - support vector learning. MIT Press, Cambridge.

Jha, Girish Nath; Hellan, Lars; Beermann, Dorothee; Singh, Srishti; Behera, Pitambar; & Banerjee, Esha. 2014. Indian languages on the TypeCraft platform–The case of Hindi and Odia. In *Proceedings of WILDRE-2014 Reykjavik, Iceland*, 84-90.

Kaur, Darvinder & Gupta, Vishal. 2010. A survey of named entity recognition in english and other indian languages. International Journal of Computer Science, 7:6.

Kaur, Darvinder & Gupta, Vishal. 2012. Name entity recognition for Punjabi language. International Journal of Computer Science and Information Technology & Security 2:3.

Nayan, Animesh; Rao, B. Ravi Kiran; Singh, Pawandeep; Sanyal, Sudip; & Sanyal, Ratna. 2008. Named entity recognition for Indian languages. In IJCNLP, 97-104.

Ojha, Atul Kumar; Behera, Pitambar; Singh, Srishti; & Jha, Girish Nath. 2015. Training & evaluation of POS taggers in Indo-Aryan languages: A case of Hindi, Odia and Bhojpuri. In *Proceedings of LTC-2015, Poland*, 524-529.

Saha, Sujan Kumar; Chatterji, Sanjay; Dandapat, Sandipan; Sarkar, Sudeshna; & Mitra, Pabitra. 2008. A hybrid approach for named entity recognition in Indian languages. In Proceedings of the IJCNLP-08 Workshop on NER for South and South East Asian languages, 17-24.

Saha, Sujan Kumar; Sarkar, Sudeshna; & Mitra, Pabitra. 2008. Gazetteer preparation for named entity recognition in Indian languages. IJCNLP. 9-16.

Sasidhar, B.; Yohan, P. M.; Vinaya Babu, A.; & Govardhan, A. 2011. A survey on Named Entity Recognition in Indian languages with particular reference to Telugu. IJCSI International Journal of Computer Science Issues, 8:2.

Srivastava, S.; Sanglikar, M.; & Kothari, D.C. 2011. Named Entity Recognition system for Hindi language: A hybrid approach. International Journal of Computational Linguistics (IJCL), 2:1.

Kaur, Amandeep; Josan, Gurpreet S; & Kaur, Jagroop. Named entity recognition for Punjabi: A Conditional Random Field approach. In the proceedings of 7th International Conference on Natural Language Processing, Macmillan Publishers, India.

Shishtla, Praneeth M; Gali, Karthik; Pingali, Prasad; & Varma, Vasudeva. 2008. Experiments in Telugu NER: A Conditional Random Field approach. In the proceedings of the IJCNLP-08 Workshop on NER for South and South East Asian Languages, 105-110.

Bikel, Daniel M.; Miller, Scott; Schwartz, Richard; &Weischedel, Ralph. 1997. Nymble: A high performance learning name-finder. In Proceedings of the Fifth Conference on Applied Natural Language Processing, 194–201.

Borthwick, Andrew. 1999. A Maximum Entropy Approach to named entity recognition. Computer Science Department, New York University, Ph.D. thesis.

Cucerzan, Silviu & Yarowsky, David. 1999. Language independent named entity recognition combining morphological and contextual evidence. In Proceedings of the Joint SIGDAT Conference on EMNLP and VLC 1999, 90–99.

Krishnarao, Awaghad Ashish; Gahlot, Himanshu; Srinet, Amit; & Kushwaha, D. S. 2009. A comparison of performance of sequential learning algorithms on the task of Named Entity Recognition for Indian languages. In the proceedings of 9th International Conference on Computer Science. Baton Rouge, LA, USA, 123-132.

Gali, Karthik; Surana, Harshit; Vaidya, Ashwini; Shishtla, Praneeth; & Sharma, Dipti Misra. 2008. Aggregating machine learning and rule based heuristics for named entity recognition. In the proceedings of the IJCNLP-08 Workshop on NER for South and South East Asian Languages, Hyderabad, India, 25-32.

Ekbal, Asif & Bandyopadhyay, Sivaji. 2007. Lexical pattern learning from corpus data for named entity recognition. In Proceedings of International Conference on Natural Language Processing (ICON).

Ekbal, Asif & Bandyopadhyay, Sivaji. Bengali Named Entity Recognition using Support Vector Machine. In the proceedings of the IJCNLP-08 workshop on NER for South and South East Asian Languages, Hyderabad, India, 51-58.

Grishman, Ralph. 1995. The New York University system muc-6 or where's the syntax? In Proceedings of the Sixth Message Understanding Conference.

Kumar N. & Bhattacharyya Pushpak. 2006. Named Entity Recognition in Hindi using MEMM. In Technical Report, IIT Bombay, India.

Li, Wei & Mccallum, Andrew. 2004. Rapid development of Hindi named entity recognition using Conditional Random Fields and feature induction. In ACM Transactions on Computational Logic.

Srihari, R.; Niu, C.; & Li, Wei. 2000. A Hybrid Approach for Named Entity and Sub-Type Tagging. In Proceedings of the sixth conference on Applied natural language processing.

Wakao T.; Gaizauskas, R; & Wilks, Y. 1996. Evaluation of an algorithm for the recognition and classification of proper names. In Proceedings of COLING-96.

INDEFINITE PRONOUNS IN ASSAMESE

Pushpa Renu Bhattacharyya

The study describes the indefinite pronouns in Assamese with its complexity and unique features and put emphasis on the morphological, semantic and syntactic properties. The term pronoun is used in the sense of the nominal expression and the substitution in the nominal slot only.

Keywords: indefinite, morphology, syntax, semantics.

1 Introduction

The present study is a modest attempt at description of the extensively used form and function of indefinite pronouns in Assamese a New Indo-Aryan language recognized by the constitution of India as one of the official languages spoken in Assam, situated in the Bhrahmaputra valley of North East India spreading an area of 78,438.00 square kilometres. According to the census report of 2011, the population of Assam is 31,169,272 the total numbers of native speakers of Assamese being 16.8 million, while the language is spoken by over 20 million people belonging to heterogeneous speech communities living together. Assamese is agglutinative with a SOV word order and nominative-accusative case alignment system with subject-verb agreement.

1.1 Pronouns in Assamese

The term pronoun is used here as pro-nouns i.e., grammatical items that can replace nouns and noun phrases. It must be mentioned that in Assamese, epithets are used as post nominal modifiers of the head proper nouns and of the second and third person head pronouns. Pronouns in Assamese are free forms. They are pronounced fully and can function deictically as well as anaphorically. They can also be suffixed with various morphemes like classifiers, plural markers, case markers, particles, etc. On the basis of their morphological representations and semantic features the pronouns may be identified and classified into different subclasses accordingly. In terms of their discourse function and frequency of use all the subclasses of pronouns can be grouped into two distinct major

subtypes - central pronouns and peripheral pronouns.

1.1.1 Central pronouns

The central pronouns consist of five subclasses:
i. Personal pronouns
ii. Demonstrative pronouns
iii. Relative pronouns
iv. Interrogative pronouns
v. Anaphoric pronouns

1.1.2 Peripheral pronouns

The peripheral pronouns consist of three subclasses:
i. Indefinite pronouns
ii. Universal pronouns
iii. Miscellaneous pronouns

The various forms of pronominals[1] in Assamese exhibit idiosyncratic morphological patterns and syntactic behaviour inherent to those specific subclasses.

2 Indefinite pronouns

An indefinite pronoun is used to refer to unknown and unidentified persons or things exhibit human vs. non-human or inanimate distinctions inherently. The pronouns belonging to this group are in a logical sense quantitative that express various degrees of indefiniteness. The label indefiniteness is used to refer to a referent which is different from, but the same kind of entity as the one referred to before. The distribution of indefinite pronouns in syntax is highly complex with semantics and pragmatics features. They have both positive[2] and negative variants. The positive indefinite pronouns are specific[3] and non-specific while the negative ones are non-specific only. Both the types of indefinite pronouns exhibit human vs. non-human or inanimate distinctions.

2.1 Positive indefinite pronouns

The positive non-specific indefinite pronouns encoding human and non-human or inanimate referents are simple and derived. The derived has compound, complex and reduplicated variants. Reduplication can be either full or partial or

modified. Table 1 displays the subtypes of positive non-specific indefinite pronouns.

Table 1: Positive non-specific indefinite pronouns

Forms	Human /Inanimate
Simple	keu / kʰenʊ / kʊnʊ 'someone' kisu 'something [-AN]'
Derived	Compound (Human/Inanimate)
	zi-kʊnʊ 'anyone', dui-eta /dui-sarita 'a few [-AN]', dui-sari / dui-ek 'a few people'
	Complex (Human/Inanimate)
	ḅale-man 'many', kisu-man 'some', zɔn–dijek 'a few people' elek-pelek 'so and so (PL)' egal-man 'many [-AN]'
	Partially reduplicated (Human/Animate)
	zei-ɦei 'anyone'
	Fully reduplicated (Human/Inanimate)
	kʊnʊ-kʊnʊ 'someone (PL)' kisu-kisu 'something(PL)[-AN]'

The types of simple indefinite pronouns have simple stems. The compounds are combinations of two free forms, e.g., two numerals dui-ek/dui-sari 'a few' or an interrogative pronoun and a relative pronoun as exemplified by zi-kʊnʊ 'anyone' respectively. A complex indefinite pronoun combines a bound form with a free form as in ḅale-man or egal-man 'many', where the bound suffix -man is used to imply approximation. A complex form may be a combination of two bound forms as in elek-pelek 'so and so'. In case of partial reduplication process two pronominal forms that belong to

different subtypes combine together as in zei-ɦei 'anyone', zi-ti or za-ta 'whatever' where the second constituent is modified. The fully reduplicated pronouns take the same simple stem as the second element as kʊnʊ-kʊnʊ 'some'.

2.2 Negative indefinite pronouns

The negative indefinite pronouns express non-existence. The negative non-specific indefinite pronouns having human and inanimate forms occur with negative verbs[4] only. Table 2 represents negative indefinite pronouns in Assamese.

Table 2: Negative indefinite pronouns

Forms	Human /Inanimate
Simple	kʊnʊ 'nobody keʊ 'none' ekʊ 'nothing'
Compound	keʊ-kisu 'nobody' ekʊ-eta 'nothing'

As seen in Table 2, kʊnʊ 'someone' is used as a positive specific indefinite pronoun as a negative non-specific indefinite pronoun as well. It has another variant keʊ 'none'. The compound forms are combinations of a simple indefinite pronoun with another pronoun or a numeral. However, the indefinites that function as nominal modifiers are not pronouns, e.g., in kʊnʊ-zɔn manuh 'some persons', kʊnʊ-zɔn is a nominal modifier, but in kʊnʊ-zɔn 'someone' in isolation is an indefinite pronoun.

3 Dependents of indefinite pronouns

Adjectives may occur with indefinite pronouns in highly context bound utterances. The expression in (1) exemplifies an indefinite pronoun taking an adjective as a pre-head modifier.

(1) [nɔtun] kiba
     new   something
     'something new'.

The following exemplifies a compound indefinite pronoun as head taking an adjective as a pre-modifier.

(2) [roŋin]   kiba-ɛta
    colourful   something
    'something colourful.'

Indefinite pronouns occur with post modifiers also. In the following example an adjective functioning as a modifier occurs as a post-head dependent.

(3) kʊnʊba [murkhɔ]
    some   idiot
    'some idiots.'

The expression in (4) provides an evidence for the fact that an indefinite pronoun functions as head that takes an indefinite quantifier as its modifier.

(4) kʊnʊba   [kei-zɔni   -man]
    IND   QUAN-CL(SG:F)-APPROX
    'a few girls.'

The least frequent type of dependent of pronouns can be an appositive as exemplified by the following with an indefinite pronoun as the head that takes a proper noun as its dependent.

(5) [kʊnʊba] [benu bɔrua]
    someone   Benu Barua
    'some Benu Barua.'

The pronouns either personal or other subtypes take quantifiers (i.e., numeral plus classifier) as post-head modifiers.

(6) ɦihÕt [du-ta]
    3_3 PL   two-CL
     'both of them.'

(7) kʊnʊba [ɛ-zɔn]
    IND   one-CL
    'some one.'

The above examples show that indefinite pronouns may take adjective (both pre and post head positions), nominal appositive and quantifier as pre modifiers.

4 Morphophonemics and morphology

The positive non-specific type of indefinite pronoun encodes human referents with direct stems- kʊnʊ, keu and kʰenʊ 'someone' and non-human referent kisu 'something [-AN]'only. These indefinites can occur with both positive and negative verbs. The first form *kʊnʊ* occurs as

subject only, while all the others can function in different case marked positions. e.g.:

(8) kʊnʊ-e   nas-is -e   kʊnʊ-e
    IND-NOM   dance-IPFV-3   IND-NOM
    ga-is-e
    sing-IPFV-3
    'Someone is dancing, someone is singing.'

(9) kʰenʊ -k   dekʰ-i   besi
    IND-ACC   see-NF   more
    utsahi   no-ho-ba
    enthusiastic   NEG-happen -2_2
    'Don't be too enthusiastic by seeing others.'

Some derived composite forms of positive non-specific indefinite pronouns are exemplified below.

(10) zikʊnʊ-e   kam-tʊ
     IND-NOM   work-CL
     kor-ibɔ   par-e
     do-NF   can- 3
     'Anyone can do the work.'

(11) ɓaleman-e   teʊ-r   kɔtʰa-t
     IND-NOM 3_2 SG-GEN   word–LOC
     ɦɔnmɔti   zɔna-l-e
     agreement   express-PST-3
     'Many agreed with him / her.'

One of the reduplicated positive non-specific indefinite pronouns shown in Table 2 exhibits a specific genitive construction, in which the initial constituent suffixed with the genitive case functions as the possessor and the final constituent as the head in a possessional construction as in elek-ɔr pelek 'so and so (PL)' used to refer to some distant indefinite referents. It can be used in both subject and object positions with relevant cases.

(12) elek-ɔr   pelek-ɔloi   iman
     IND-GEN   IND-DAT   so
     sinta   no-kor-ib-a
     worry   NEG-do-FUT-2_2
     'Don't bother much about so and so.'

The positive non-specific indefinite type has a unique partially and modified reduplicated variant with distinctions of humanness and animacy used to imply 'anyone' or 'anything.' The direct stems

used for [+HUM] are zi-ħi and zei-ħei and its indirect stems are za-ta. It has a reduced variant without the initial phonemes /z-/ and /t-/ or /ħ-/ as in genitive marked a-r ta-r and accusative marked a-k ta-k. The [-AN] type has different variants that can be used to encode direct object only. Its indirect stems are zih-tih as illustrated in Table 3.

Table 3: Reduplicated indefinite pronouns

Case	[+Human]	[-Animate]
Nominative	zi-ħi-(e), zei-e - ħei-e	zi-ti, za-ta, zi-ki
Accusative	za-k ta-k , a-k ta-k	zih-ɔk(e) tih-ɔk(e), ziti, za ta
Genitive	za-r ta-r, zeiħei-r	a-r ta-r, zih-ɔr tih-ɔr
Dative	zeiħei -loi, zata –loi	ziti-loi, zata-loi

As seen from Table 3, in case of some reduplicated pronouns of this type, the same case is suffixed to both the constituents or both the constituents remain unmarked; while in others just one case is suffixed to the reduplicated constituent as a whole as exemplified.

(13) zei-e   ħei-e   ene   kam
     IND-NOM   IND-NOM   like   work
     kor-ibɔ   nʋ-(p)ar-e
     do-NF   NEG-can-3
     'Anybody cannot do such a work.'

(14) zi –ti-∅   no-ko-ba
     IND.ACC   NEG- say-2_2
     'Do not speak nonsense.'

(15) a-k   ta-k   ħuɖ - i
     IND-ACC   IND-ACC   ask-NF
     tʋma-r   gɔr-tʋ   ulia-l-ʋ
     2_2 SG-GEN house-CL   find-PST-1
     'I have managed to find your house by asking one and another.'

Before giving a description of the positive specific type, it is a necessary prerequisite to illustrate the negative type, as both are compositionally related.

The negative non–specific indefinite pronoun has both [+HUM] and [-HUM] variants as displayed in the Table 4 below.

Table 4: Negative non-specific indefinite pronouns

Case	[+Human]	[-Human]
Nominative	kʋnʋ(e) /kʋneʋ / keu-e	ekʋ(e)
Accusative	kakʋ	ekʋ
Genitive	karʋ	-
Dative	karʋ-loi	ekʋ-loi

As presented in Table 4, the [+human] negative indefinite pronoun has three ʋ-ending direct stems. They are- kʋnʋ-, kakʋ- and karʋ- 'nobody' / 'none'. The first is used to refer to subject, the second encodes direct object, while the third encodes possessor. As shown in the Table 4, the non-human indefinite pronoun ekʋ 'nothing' shares the same direct stem for encoding all functions. Both the pronouns thus formed can be optionally be suffixed with the overt nominative marker depending on the transitivity of the verb. While there is no form for the genitive case for the non-human category, the accusative of both are unmarked for case, but the dative of both are overtly marked. The following exemplifies negative indefinite pronouns as subjects.

(16) azi-loi-ke   ta-k
     today-DAT-EMPH 3_3SG:M- ACC
      keu-e   dekh-a   n-a-e
     IND-NOM   see-NF   NEG-be 3
     'No one has seen him till today.'

(17) kʋneʋ- ɸ   ta-k
     IND-NOM 3_3SG:M- ACC
     na-mat-il-e
     NEG-call-PST-3
     'None had invited him.'

The examples in the following show its use as the object.

(18) kɔtha-tʋ   kakʋ- ɸ   no-ko-ba
     fact -CL   IND. ACC   NEG-say-FUT 2_2
     'Don't divulge the fact to anyone.'

(19) ta-k   ekʊ- ɸ   no-ko-ba
3₃SG:M-ACC IND.ACC NEG-say-FUT2₂
'Don't tell him anything.'

The following is an example of compound negative non-specific indefinite pronoun.

(20) ta-r   keʊ-kisu n-a-e
3₃SG:M-GEN   IND   NEG-be 3
'He has no one (i.e., kith and kin).'

Negative indefinite co-occurs only with negativised verbs as substantiated by the ungrammaticality of following sentences.

(21) * gɔr-ɔt   kʊnʊ as-e
house-LOC IND   be-3

(22) * kɔtʰa-tʊ   kakʊ   ko-ba
fact -CL   IND   say –FUT 2₂

Of the [+HUM] positive specific subtype with the specificity marker –ba the direct variant kʊnʊba- takes the nominative case and the oblique variant karʊba takes all other case markers as in the following.

Table 5: Positive specific indefinite pronouns

Stem	Case	Word form
kʊnʊba- karʊba-	nominative accusative genitive dative	kʊnʊba-(e) karʊba-k karʊba-r karʊba-loi

The following are examples of positive specific indefinite pronouns occurring in various syntactic positions.

(23) kʊnʊba-e   karʊba-k
IND-NOM   IND-ACC
mar-is-e
beat-IPFV-3
'Someone is beating some other.'

(24) kʊnʊba- ɸ   ah-is-e
IND-NOM   come-IPFV-3
'Someone is coming.'

(25) kali   karʊba-r   gɔr-ɔt
yesterday   IND-GEN   house-LOC

suri   ho-is-e
theft   happen-IPFV-3
'There had been a theft yesterday at someone's   home.'

The positive specific indefinite pronouns ki- and kih- used to refer to non-human and /or animate, non-animate referents are complex forms derived by the insertion of the specificity marker –(ɔ)ba by the process of anaptyxis. It has two direct variants, one is kiba-, which does not take the overt nominative case and the other is kihɔba- 'something', formed by insertion of the vowel ɔ, followed by the specificity marker ba, which takes all case markers as shown by Table 6.

Table 6: [–HUM/±AN] Positive specific indefinite pronouns

Direct stem	Case	Word form
ki-/kih-	nominative	kiba / kihɔba-e
	accusative	kiba / kihɔba-k, ziba
	genitive	kihɔba-r
	dative	kihɔba-loi

The sentences in (26) and (27) exemplify the use of positive specific indefinite pronouns.

(26) kiba- ɸ   ko-ba neki
IND.ACC say-2₂ QF
Do you want to say something?'

(27) kihɔba-e   mat-is-e
IND-NOM   call-IPFV-3
'Something is producing a sound.'

These ba- ending indefinite pronouns need to be distinguished from the interrogative-indefinite pronouns with the dubitative particle -ba used to express the speaker's sense of doubt or anxiety interspersed with the question. Phonologically, the question word carries the rising intonation, while the dubitative particle is marked by a falling intonation. It has a limited number of contrastive forms for both human and non-human referents. Table 7 represents those pronouns.

Table 7: Interrogative-indefinite pronoun

Case	[+Human]	[-Human]
Nominative	kʊn-ba	ki-ba
Accusative	kak-ba	kihɔk-ba
Genitive	kar-ba	kihɔr-ba
Dative	kaloi-ba	kihɔloi–ba

The use of these pronouns in various case marked positions are shown below.

(28) kʊn-ba - ɸ   ah-is-il
    INT-IND-NOM   come -IPFV-PST3
    'Who might have come?'

(29) ħi- ɸ   ki-ba- ɸ
    3₃SG:M- NOM   INT-IND-ACC
    an-is-e
    bring-IPFV-3
    'What might he have brought?'

The bound form–(ɔ)ba is used as a specific marker suffixed to question words or interrogative pronouns to derive the complex forms of positive specific indefinite pronouns like kʊnʊ-ba 'someone', kih-ɔba 'something', ki-ba 'something' and relative pronoun zi-ba 'whichever / whatever [-AN]', encoding human, non-human, animate and inanimate referents respectively. Some of them exhibit compound forms, e.g., kiba-eta 'something', kiba-kibi 'somethings' etc.

However, it has another ba- ending variant used as interrogative-indefinite pronoun, where a case marked interrogative pronoun is suffixed with –ba as a dubitative particle to express the speaker's sense of doubt or anxiety interspersed with the question. Phonologically, the question word carries the rising intonation, while the dubitative particle is marked by a falling intonation. It has a limited number of contrastive forms for both human (e.g., kʊn-ba, kak-ba, kar-ba) and non-human (e.g., ki-ba, kihɔk-ba, kihɔr-ba) referents.

However, the same form suffixed to the indefinite quantifier *kei* as *kei-ba* followed by classifiers may be used to imply 'quite a few' (Chowdhary 2012: 272) with reference to indefinite referents

of various types of animacy, e.g., kei-ba-zɔn (human), kei-ba-ta (human / non-human / inanimate).

5 Marking of number

Number distinction has no grammatical bearing on the language. The morphological processes associated in the formation of plural forms of the various subclasses of pronouns are illustrated below.

Except few, most of the indefinite pronouns are inherently neutral in exhibiting number differentiations. The morphological processes associated in the formation of derived plural indefinite pronouns may be summed up as in the following:

i. Suffixation of plural markers

kʊnʊ'some (SG)'>kʊnʊ - bʊr / -bilak 'some (PL)'

ii. Compounding

zɔn-oik 'one person'> zɔn-diek 'more than one', kiba-eta 'something [-HUM]'>kiba-kibi 'something (PL)[-AN]';

kʊnʊ'some (SG)'>kʊnʊ-kʊnʊ'some (PL)'. elek-pelek 'so and so(PL)';

iii. Reduplication:

kiba'something (SG)'>kiba-kibi'somethings', kʊnʊ'some (SG)'>kʊnʊ-kʊnʊ 'some', kʊnʊba 'some (SG)'>kʊnʊba kʊnʊba 'some (PL)'

6 Marking of gender

In case of some indefinite pronouns gender is distinguished on the basis of whether the masculine classifier -zɔn, -tʊ or the feminine classifier -zɔni is used as its final constituent, e.g., kʊn-tu 'which one (M)' > kʊn-zɔni 'which one (F)'; kʊnʊ-zɔn 'someone (M)'> kʊnʊ-zɔni 'someone (F)' respectively.

Gender in pronouns of Assamese is found to be only restrictively grammatical.

7 Semantic features of indefinite pronouns

The semantic feature of indefiniteness characterizes indefinite pronouns. They may encode positivity or negativity. The positive indefinite pronouns may again imply specificity or non-specificity. The semantic distinctions of animacy/inanimacy and humanness/non-humanness are the additional salient features of indefinite pronouns. The final element –ʋ, signalling indefiniteness is shared by most of the positive non-specific indefinite pronouns like kʋnʋ, keu, khenʋ, kisu 'someone' as well as negative non-specific indefinite pronouns karʋ, kakʋ 'none' and ekʋ 'nothing' etc.

Specificity is associated with -(ɔ)ba- ending indefinite pronouns as in kʋnʋba-, karʋba- 'someone', kiba 'something', kihɔba 'something'.

The positive non-specific indefinite pronouns suffixed with classifiers indicate sex distinctions as in kʋnʋzɔn 'someone (SG:M)', kʋnʋzɔni 'someone (SG:F).'

Reduplicated indefinite pronouns are used to mark plurality as in zeiħei 'anyone', ziti/zata 'anything', zikʋnʋ 'any one', kʋnʋkʋnʋ 'some.'

The meaning differences can be seen in NPs with indefinite pronouns as the head with or without dependents. For example, kiba 'something' or kʋnʋba 'someone' can have a specific sense when it occurs with a quantifier as it's dependent as in kiba eta 'something' or kʋnʋba ezɔn 'someone.'

8 Pragmatics of indefinite pronouns

8.1 Indirectness and politeness

Indefinite pronouns are very deeply related with the concept of politeness. It has been argued that there is a correlation between indirectness and politeness. In general not to address another person too directly is considered to be a very general politeness strategy. For example, the subject position of the imperative sentence which is canonically characterized by an ellipse subject can be overtly filled up by an indefinite pronoun in place of the 2nd person pronoun.

(30) kʋnʋba -   zʋ-a   aru
      IND -NOM   go-IMP.2₂   and
      sɔkidar - zɔn   - ɔk   mat-i
      chowkidar –CL(SG:M)– ACC   call-NF
       an-a
      bring-IMP.2₂
      'Someone go and call the chowkidar.'

The opposite also holds true when in a specific context of discourse, the positive specific indefinite pronoun is used to make sarcastic remarks or utterances with a sense of mirth towards the addressee, e.g.,

(31) kʋnʋba-e   ama – k   mɔt-a
      IND-NOM   1PL-ACC   call-NF
      n-a-e
      NEG-be-3
      'Someone has not talked to us.'

This may be interpreted as 'Why are you not talking to me?'

The following is a conversation between a mother and her child expressing the small steps of politeness, where the mother indirectly accuses the child and the child tries to exonerate himself/ herself from the blame.

(32) Mother
      kʋnʋba-e   kɔl   kʰa – l-e
      IND-NOM   banana   eat-PST-3
      'Someone has eaten banana(s).'

      Child
      mɔe - ɸ   kʰʋ-a   nae
      1SG –NOM   eat-NF   NEG
      'I have not.'

The use of the positive specific indefinite pronoun renders a polite overtone to a question by turning it into an indirect one, as exemplified by the following pairs of utterances with the interrogative pronoun ki and the indefinite pronoun kiba.

(33) a. ki   ho-l
         INT   happen-PST.3
         'What happened?'

(33) b. kiba   ho-l   neki
   IND   happen-PST.3   QF
   'Has anything happened?'

The same is the case in the following pairs of utterances, where the use of the positive specific indefinite pronoun in sentence (b) encodes a polite enquiry about the stock of food in the house, in contrast to (a) with an interrogative pronoun which sounds more like a demand to know about it.

(34) a. kʰa-boloi   ki   as-e
   eat-NF   INT   have-PRES.3
   'What is there to eat?'

(34) b. kʰa-boloi kiba
   eat-NF   IND

   as-e   neki
   have -PRES.3   QF
   'Is there something to eat?'

8.2   Generic Sense

The positive non-specific indefinite pronouns are the best examples of generic interpretations, e.g.,

(35) zikʊnʊ-e   kam-tʊ   kor-ibɔ   par-e
   IND-NOM work-CL do-NF   can-3
   'Anyone can do the work.'

8.3   Special Sense

The positive non-specific indefinite pronouns suffixed with classifiers indicate sex distinctions as in kʊnʊ-zɔn 'someone (SG:M)', kʊnʊ-zɔni 'someone (SG:F)'.

(36) kʊnʊ- zɔni-e   nas-is -e
   IND-CL(SG:F)-NOM   dance-IPFV-3
   kʊnʊ- zɔni-e   ga-is-e
   IND- CL(SG:F)-N   sing-IPFV-3

'Someone (F) is dancing, someone (F) is singing.'

9 Conclusions

The foregoing was an attempt to present the main features of indefinite pronouns in Assamese in terms of the forms and functions. The study shows that indefinite pronouns play an important role in the language.

References

Aldridge, Maurice V. 1982. *English quantifiers: A study of quantifying expressions in linguistic science and modern English usage.* Amersham: Avebury.

Bach, E. 1968. Nouns and noun phrases. In E. Bach, and Robert T. Harms (eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart, &Winston, 91-124

Bakker,   C.L.   1970.   Double   negatives. *Linguistic Inquiry*, 1:169-86.

Bhat, D. N. S. 1993. *Indefiniteness and Indeterminacy*.   MS,   Mysore,   Central Institute of Indian Languages

Chowdhary,   R.C.   2012.   On   Classifiers   in Asamiya. In Hyslop G., S. Morey and M.W. Post (eds.) *North East Indian Linguistics,*Vol.4, Cambridge University Press India Pvt. Ltd.

Online References

Assamese Language, Web Retrieved from
https://en.wikipedia.org/wiki/Assamese_language
on 7th August, 2017

---

[1] The term pronominal is used as replacer of nominal constructions.

[2] The term positive is used to describe presence   while negative is used to describe absence of an indefinite quantity.

[3] The term specific refers to a particular instance of a class of referents and non-specific refers to the whole class of entities.

[4] Examples of negative indefinite pronouns occurring with negative verbs are shown in sub section 4 (examples 16-20).

ASPIRATION IN NEPALI

Krishna Prasad Chalise

This experiment studies the effect of aspiration on PVD, CD, ACT and SA of the Nepali plosives and finds that the voiced aspirates (also called breathy voiced) share the features of both the voiced plosives and aspirated plosives. So, they are voiced aspirates phonetically. To define aspiration in terms of voicing lag is inappropriate because it is found in both the voiceless and voiced plosives in Nepali.

1 Introduction

Phonologically, Nepali has 16 plosive sounds with four-way contrast in terms of phonation: voiceless vs. voiced and unaspirated vs. aspirated, and four way contrast in the places of articulation: bilabial, dental, retroflex and velar. As in Hindi (Lisker and Abramson 1964, Ohala and Ohala 1972, Benguerel and Bhatia 1980 and Dutta 2007) and Bengali (Mikuteit and Reetz, 2007), whether the voiced aspirates in Nepali are phonetically aspirated plosives or they represent a distinct mode of phonationhas been a matter of debate (Poon and Mateer1985, Pokharel 1989 and Chalise 2017).

In the Sanskrit phonetic literature the voiced aspirates have been identified as the combined articulation of *nāda* (voicing) and *svāsa* (aspiration) and clearly described the distinction between the two phonetic features (Dutta 2007:3). In contrast to the Sanskrit phonetic tradition, Ladefoged (1975: 127) regards voiced aspirate to be a distinct mode of phonation. He claims that the voiced aspirates are phonetically distinct sounds because 'they are neither voiced (in the sense of having regular vibrations of the vocal cords) nor aspirated (in the sense of having a period of voicelessness) during and after the release of the closure'. So he uses the terms breathy voice or murmur to refer to the voice aspirates. The idea of Ladefoged has been established as a standard view about the voice aspirates in phonetics.

But there are a number of phoneticians who support the idea that voiced aspirates are the combination of voice and aspiration. As cited in Dutta (2007:7) 'Lombardi (1994) argues that the voiced aspirates in Hindi and other languages are both voiced and aspirated'. Her argument is based on the findings of Dixit (1989), Yadav (1984), Ingemann and Yadav (1978) and Kagaya and Hirose (1975). Ingemann and Yadav (1978) claim that the concept of Ladefoged (1975) has made the pattern of plosives in four-category languages asymmetrical and counter intuitive; and it has created problems with the historical description of the sounds in question.

But Ladefoged and Johnson (2011) assert the voiced aspirates to be voiced sounds and the term 'breathy voiced' used to refer to the class itself asserts them to be voiced sounds.It shows that there is no controversy regarding the voicing of the voiced aspirates but the controversy is whether they are aspirated or not. So the foundation of the controversy lies on how we define aspiration.

There are some issues to be dealt with while we try to understand what aspiration is. The first issue is whether aspiration is the feature of any class of sounds, only the plosive sounds or only the voiceless plosive sounds. The second issue is whether aspiration is phonetically noise or voicelessness. Similarly, which phase of sound production, onset, hold or release, it belongs to.

Laver (1994: 348) regards aspiration to be the sole property of voiceless plosive. He states that 'aspiration is a feature which can manifest a co-ordinatory relationship between a voiceless segment and a following voiced at the leading edge of a syllable'. Similarly, a large number of phoneticians regard it to be the property of voiceless plosives and a longer duration of voicelessness or voicing lag (Voice Onset Time) after the release of a plosive indicates an aspirated one and a short voicing lag indicates an unaspirated one (Ladefoged and Johnson 2011:305; Raphel, Borden and Harris 2011:134; Reetz and Jongman 2011:101). From the above literature we understand that aspiration is a release feature of a voiceless plosive. According to this concept, the voiceless unaspirates and aspirates

are characterized by short and long voicing lag respectively which is technically described using the term Voice Onset Time (VOT).

This concept presupposes that voicing lag is found only in the voiceless plosives but in Nepali voicing lag is found in voiced plosives as well and this phenomenon is found in other languages like Bengali (Mikuteit and Reetz, 2007) and Georgian (Vicenik, 2008), too. In Georgian, aspirated stops show longest voicing lag and voiced stops show shortest voicing lag. So voicing lag is found in both voiced and voiceless plosives in the languages. The concept is contradictory because if aspiration is only the feature of the voiceless plosives, voiced plosives should not have voicing lag and if voicing lag is aspiration, it is possible with the voiced plosives, too. So the concept of VOT is contradictory in itself and it does not seem to be logical for the four category languages like Nepali which has already been accepted by Lisker and Abramsom (1964) and Poon and Mateer (1985).

Similarly, aspiration does not seem to be only the feature of plosives as well as their release. In Dzonkha, the national language of Bhutan has contrastive pre-aspirated sonorants as in ləp 'hand' vs. hləp 'more' (Watters 2002:9) which shows that aspiration is a property of a sonorant consonant in the onset position, too.

The VOT model uses the terms 'voicing lag/voicelessness and 'noise' alternatively to refer to aspiration. But voicelessness does not necessarily mean noise because a silence is also a state of voicelessness. So whether aspiration is a type of noise or simply the state of voicelessness is a matter to discuss. Some phoneticians regard aspiration to be a noise produced as a result of glottal and supraglottal friction. Harrington (2010:55) states that 'during the release of a stop the air is pushed out of the vocal tract explosively and this brief period of high airflow may result in aspiration noise'. Similarly, he Harrington (2010:103) further clarifies that 'aspiration is the result of noise source at the glottis that may produce energy below 1 kHz'.

There are several debates regarding the different aspects of aspiration, so in this experiment, based on the Nepali data, I have tried to find out and compare the acoustic features of the corresponding aspirated and unaspirated plosives, and voiced and voiceless aspirated plosives based on their effects on the preceding vowel, closure time features and release features.

2 Methodology

2.1 Method of data collection

The words presented in Table 1 consists all the target sounds in [iCi] and [aCa]. Every word was embedded in a carrier sentence as: X, I said X (where X is the target word) where the word is uttered as a single word for the first time and a part of the utterance for the second time, and the speakers were asked to utter for three times. Every utterance was followed by a pause so that the speaker could produce each utterance with equal comfort. For the purpose of analysis the word produced as a word was selected. Altogether, there were [16 (number of plosives) × 2 (number of environments) × 3 (one word was uttered for three times) × 6 (number of speakers)] 576 tokens for analysis.

Table 1: The word list

[k]	[ʈ]	[t]	[p]
[ʈiki, kaka]	[piʈi, paʈa]	[riti, tata]	[pipi, papa]
[kʰ]	[ʈʰ]	[tʰ]	[pʰ]
[tikʰi, kakʰa]	[chiʈʰi, paʈʰa]	[tithi, gatha]	[pʰipʰi, lapha]
[g]	[ɖ]	[d]	[b]
[bigi, jaga]	[piɖi, paɖa]	[didi, dada]	[bibi, baba]
[gʱ]	[ɖʱ]	[dʱ]	[bʱ]
[gʱigʱi, agʱat]	[piɖʱi, gaɖʱa]	[dʱidʱi, adʱa]	[ʈibʱi, abʱar]

The utterances were recorded using Sony ECM-MS908C Electret Condenser Microphone and EDIROL, R09HR audio recorder maintaining a distance of 10-12 inches between the microphone and the mouth of the speaker in waveform files with 44000 Hz audio sample rate, 1411 bit rate and 24-bit resolution.

2.2 The speakers

Six fluent native speakers of Nepali, three males and three females, with normal speech capacity were recruited for the experiment. The speakers have been included from different age groups as presented in Table 2.

Table 2: The sample of the speakers

S. N.	Age group	Gender	
1	21-30	male [DA]	female [JA]
2	31-40	male [HR]	female [GY]
3	41-50	male [KR]	female [KL]

2.3 Analysis of the data

The recorded data were edited using Audacity, an audio editing software and were analyzed using PRAAT a sophisticated and widely used software for acoustic analysis. This study has focused on the effect of aspiration on the plosive itself and the preceding vowel and the following vowel. Oscillogram, spectrum and spectrogram of the sounds were used as the devices for analysis. The techniques of measurement are based on Ladefoged (2003). The statistical calculations were made using http://vassarstats.net/, a statistical calculation website.

This research has followed the segmentation model proposed by Mikuteit and Reetz (2007) for the study of the East Bengali plosive sounds. This model proposes four phases in vowel-plosive-vowel sequence as presented in Fig. 1.

Preceding vowel duration (PVD)

It is the duration of the vowel preceding the plosive. Its duration begins with the end of the preceding segment and ends at the beginning of the closure duration of the plosive.

Closure Duration (CD)

It is the duration of the hold phase of the plosive production. It begins with the end of the preceding vowel and ends at the release of the closure. The end of the preceding vowel is indicated by the sudden cessation of the high amplitude vocal fold vibration of the vowel and beginning of the silence (for voiceless) or very low amplitude buzz (for voiced). Similarly, the end of the closure is indicated by beginning of the release burst indicated by a short spike.
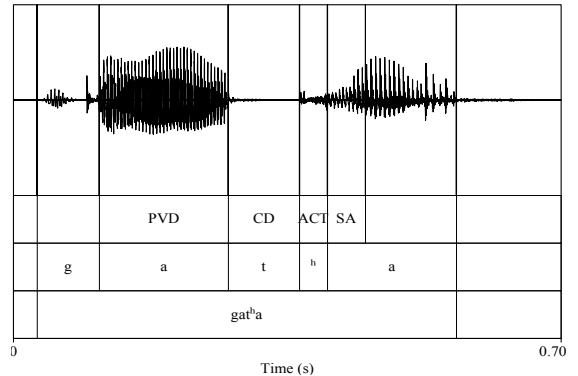
Figure 1: The phases in vowel-plosive-vowel sequence

After Closure Time (ACT)

This is the duration from the first release burst to the beginning of the first regular glottal pulses of the following vowel. The ACT is characterized by the aperiodic noise component. Itis the term used instead of VOT where VOT is the voicing lag after the burst of a voiceless plosive but ACT is used to refer to such aperiodic noise in both voiceless and voiced plosives.

Superimposed aspiration (SA)

It begins with the end of the after closure time (ACT) and ends at the decrease of end of the friction noise caused by aspiration. It is rather difficult to identify the end of SA because there is no clear demarcation to indicate the end of SA; it is a continuum. It gradually decreases from the beginning to the end of the vowel. Similarly, the nature of SA varies according to the nature of the plosive. It is lighter following an unaspirate and heavier following an aspirate. Similarly, it varies according to the voicing of a plosive as it is lighter following an unaspirated and heavier following an aspirate.The effect of the aspiration can be seen in a spectrogram as the degree of fadedness.

3 The results

3.1 Aspiration and the preceding vowel duration (PVD)

The average PVD is slightly longer for aspirates (175ms, SD (standard deviation) =40) and slightly shorter for unaspirates (172 ms, SD=33). But a one-way ANOVA indicates that the PVD does not depend on the aspiration of a plosive (F (1, 15) = 0.33, p=0.5741).The result is the same, separately, for both voiced and voiceless plosives. For voiceless aspirates the average PVD is 152 ms (SD=32) and for voiceless unaspirates the average PVD is 150 ms (SD=26) where F (1, 7) =0.06, p=0.8135. Similarly, for voiced aspirates the average PVD is 199 ms (SD=32) and for voiced unaspirates the average PVD is 194 ms (SD=25) where F (1, 7) = 0.26, p=0.6257. It suggests that PVD is independent of aspiration both in voiceless and voiced plosives. Moreover it suggests that the relation between voiceless unaspirates and voiceless aspirates is parallel to the relation between voiced unaspirates and voiced aspirates regarding the PVD.

3.2 Aspiration and closure duration (CD)

Aspiration has significant effect on the CD of a plosive as aspirates are shorter than the unaspirates in the approximate ratio of 3:4. The average CD for the aspirates is 63 ms (SD=11) and the average CD for the unaspirates is 80 ms (SD=21). A one-way ANOVA indicates that the relationship between aspiration and closure duration is statistically very significant as F (1, 15) = 27.52, p<0.0001.

The relationship between aspiration and CD of the plosives is the same,separately, for both voiced and voiceless ones. For the voiceless aspirates, the CD is 72 ms (SD=6) and for the voiceless unaspirates CD is 98 ms (SD=13) where F (1, 7) = 31.81, p = 0.0007. Likewise, the CD for voiced aspirates is 54 ms (SD=8) and for voiced unaspirates is 63 ms (SD=6) where F (1, 7) = 19.3, p=0.0031. So, the facts justify that the relation between voiceless unaspirates and aspirates is parallel to the relation between voiced unaspirates and aspirates regarding the CD.

3.3 Aspiration and After Closure Time (ACT)

Aspiration has significant effect on ACT. The ACT of the aspirates is remarkably longer than that of the unaspirates. It is comparatively far longer in the voiceless plosives than in the voiced

plosives so the comparison of the average values of the voiced and voiceless plosives will not be logical.

For the voiceless aspirates, the average ACT is 69 ms (SD=20) and for the voiceless unaspirates, it is 20 ms (SD=9) where F (1, 7) = 99.37, p<0.0001. Similarly, for the voiced aspirates, the average ACT is 7 ms (SD=3.7) and for the voiced unaspirates it is 4 ms (SD=1.7) where F (1, 7) = 3.88, p=0.0895 (which is slightly less than the required confidence level). The burst and frication period of the voiced plosives is very weak in Nepali and the voiced aspirates are produced fricatives in most of the situations. So measuring ACT in the Nepali plosives is indeed challenging. Still the relation between the voiceless aspirates and voiceless unaspirates; and voiced aspirates and voiced unaspirates have the same and parallel pattern.

3.4 Aspiration and Superimposed Aspiration (SA)

Aspiration has substantial effect on the SA of a plosive. It is longer in the aspirates and shorter in the corresponding unaspirates in the approximate ration of 2:1. The average value for the voiced aspirates is 53 ms (SD=7) and for voiced unaspirates is 24 ms (SD=8) where F (1, 7) = 133.83, p<0.0001. Similarly, the average value for the voiceless aspirates is 28 ms (SD=6) and for voiced unaspirates is 16 ms (SD=6) where F (1, 7) = 16.05, p<0.0051.

It indicates that aspiration has remarkable effect on the SA of a plosive in both voiceless and voiced plosives. Moreover it suggests that the relation between voiced aspirates and voiced unaspirates and voiceless aspirates and voiceless unaspirates is parallel regarding their SA values.

4 Discussion and conclusions

This experiment shows that the characters that voiceless unaspirated plosivesdepict are completely shared by the voiceless aspirated plosives and all the features depicted by the voiced unaspirated plosives are shared by the voiced aspirated plosives, too. So the voiced aspirated plosives make a common class with their unaspirated counterparts. Similarly, the patterns of acoustic relations between the voiceless unaspirated plosives and the voiceless

aspirated plosives are parallel to the patterns of acoustic relations between the voiced unaspirated plosives and the voiced aspirated plosives. The relation can be presented symbolically as:

voiceless unaspirated plosive : voiceless aspirated plosive = voiced unaspirated plosive : voiced aspirated plosive

The parallel phonetic features between the voiceless aspirated plosives and the voiced aspirated plosives with reference to their unaspirated counterparts justifies that voiced aspirated plosives are aspirates in reality but not a distinct mode of phonation as mentioned in the literature. So it is not logical to regard the voiced aspirates as a distinct mode of phonation and term them as 'breathy voiced' but they are the combination of voicing and aspiration.

The term 'breathy voiced' is misleading because the term 'breathy' is used to refer to the breathy sonorants and vowels which are voiced. So the breathy sonorants and vowels are breathy voiced in reality. The next point to remember is that breathiness is a hold period feature of the sounds but the term 'breathy voiced' has been used based on the after release feature of the plosive and more over based on the effect of the plosive on the following vowel, in fact, which is not the part of the plosive. In reality, the breathiness in the following vowel is found in the case of voiceless aspirates, voiceless unaspirates and voiced unaspirates, too.
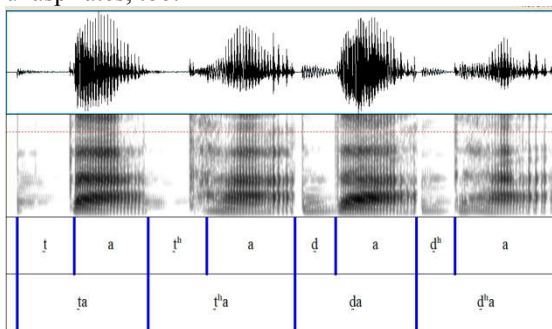
Figure 2: The breathiness in the vowel after [t̪], [t̪ʰ], [d̪] and [d̪ʰ] compared

It is a matter of degree, as in Figure 2, which can be presented in order as: voiceless unaspirate<

voiced aspirate < voiceless aspirate < voiced aspirate.

So aspiration needs to be defined to accommodate the voiceless aspirates and voiced aspirates as well. If we define aspiration as the release feature of a plosive as an extra air released with glottal and supraglottal friction causing noise, we can handle the aspiration in the four-category languages like Nepali, Hindi, etc. The release of a plosive includes transient (burst), frication (the opening period which is similar to the fricative production), and optionally aspiration (noise as a result of friction at the glottis and supraglottal region).

It would maintain the parallel terms between phonetics and phonology and we should not bear the expense of another new term.

References

Benguerel, Andre P. & Bhatia, Tej K. 1980. Hindi stop consonants: An acoustic and fiber-scopic study. *Phonetica*, 37, 134–148.

Chalise, Krishna P. 2017. Acoustic analysis of the plosives in Nepali. *Gipan*, 3.1: 41-64: Central Department of Linguistics, T. U., Kathmandu.

Dixit, R. P. 1989.Glottal gestures in Hindi plosives.*Journal of Phonetics*, 17.3, 213-237.

Dutta, Indranil. 2007. *Four-way contrast in Hindi: an acoustic study of voicing, fundamental frequency and spectral tilt*. A dissertion submitted in partial fulfillment of the requirements for the degree of doctor of philosophy in linguistics in the Graduate College of the University of Illinois. Urbana-Champaign.

Harrington, Jonathan. 2010. Acoustic Phonetics. In *The Handbook of Phonetic Sciences*. William J. Hardcastle, John Laver and Gibbon Fiona (Eds.), 81-129. Chichester: Wiley-Blackwell.

Ingemann, Frances and Yadav, Ramawatar. 1978. Voiced aspirated consonants.*Papers from the 1977 Mid-America Linguistics Conference*. Columbia.University of Missouri, pp. 337-344.

Johnson, Keith. 1997. *Acoustic and Auditory Phonetics*. Cambridge and Oxford: Blackwell Publishers.

Kagaya, R. and Hirose, H. 1975. Fiberoptic, electromyographic and acoustic analyses of Hindi stop consonants. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* (Tokyo). 9, 27-46.

Ladefoged, Peter and Keith Johnson. 2011. *A Course in Phonetics.* 6th Boston: Wadsworth Cenage Learning.

Ladefoged, Peter. 1975. *A course in phonetics.* New York: Harcourt Brace Jovanovich.

Ladefoged, Peter. 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Oxford: Blackwell publishers.

Laver, John. 1994. *Principles of phonetics*: Cambridge University Press.

Lisker, L. and Abramson, A. S. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word,* 20, 384-422.

Lombardi, Linda. 1994, *Laryngeal features and laryngeal neutralization*(Outstanding dissertations in Linguistics). Garland Publishing.

Mikuteit, Simon. & Reetz, Henning. 2007. Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language & Speech*, 50.2, 249-279.

Ohala, Manjari and Ohala, John. 1972. The problem of aspiration in Hindi phonetics. In *Annual Bulletein of Research Institute of Logopedics and Phoniatrics*, Faculty of Medicine, University of Tokyo, 6, 39-46.

Pokharel, Madhav.P. 1989. *Experimental analysis of Nepali sound system*. Ph.D. Dissertation. Deccan College, Pune.

Poon, Pamela G. and Mateer, Catherine A. 1985, A study of VOT in Nepali stop consonants, *Phonetica*42, 39-47.

Reetz, Henning, & Jongman, Allard. 2009. *Phonetics: Transcription, production, acoustics, and perception*. West Sussex, UK: Wiley-Blackwell.

Vicenik, Chad. 2008. *Acoustic study of Georgian stop*. Working Papers in phonetics, Paper1_No107 in Department of Linguistics, UCLA.

Watters, Stephen A. 2002. *The sounds and tones of five Tibetan languages of the Himalayan region*. Linguistics of the Tibeto-Burman Area, 22.1, 1-65.

Yadav, Ramawatar. 1984. Voicing and aspiration in Maithili: a fiberoptic and acoustic study. *Indian Linguistics*. 45, 1-25.

Acknowledgements

ECHO-WORD FORMATION IN BANGLA

Kuntala Ghosh Dastidar

This paper describes and analyzes echo-word formation in Bangla, one of the eastern Indo-Aryan languages of South Asia. Echo-word formation, one kind of partial reduplication, has lexical as well as syntactic implications in Bangla grammar. More specifically, this paper investigates different functions of this partial reduplication and presents its unified account in the grammatical system of Bangla.

1 Introduction

Emeneau (1956), Masica (1976) have defined the Indian subcontinent as a "Linguistic area" i.e., an area that includes languages belonging to more than one language family but sharing the linguistic features that prototypically do not belong to a specific language families because of the age long contacts. This linguistic area comprises languages primarily belonging to Indo Aryan, Dravidian, Tibeto-Burman and Austro-Asiatic language family and some language isolates. These genetically unrelated languages share some common traits as a result of their geographical proximity and language contact. Echo words (EW) or Echo Reduplication (ER) is one of such common areal features.

Bangla, an Indo-European, has been influenced by languages of other language families spoken in this linguistic area. These genetically unrelated languages have contributed to Bangla vocabulary and provided the language with some structural forms. Emeneau (1956:10) has claimed that echo words are indeed a pan-Indic trait but Indo-Aryan languages like Bangla have received it from non-Indo-Aryan languages as it has not been found in Indo-European languages.

Echo word formation is a kind of Lexical Reduplication (LR) in which the base word is duplicated (in some cases preceded also) by an echo word which is actually the partially reduplicated form of the base word. It is partially reduplicated rather repeated because of the replacement of the initial phoneme (vowel or consonant) or syllable of its base word. So, ER is basically 'Partial Lexical Reduplication'. The meaning conveyed by the echoed part of a base word is mainly 'and the like or etc.'

(1)   a. Bangla
          pʰul -ʈul
          flower-ER   "flowers and the like"

      b. Hindi
          nam-vam
          name-ER "names and the like"

2 Structure of echo words in Indian languages

Echo words are formed mainly by replacing the initial phoneme or syllable of the base words with a completely different phoneme or syllable. These replacing phonemes or syllables vary according to the language, but in a particular language it is more or less fixed and rigid. Echo word formation has been studied by Emeneau 1956, Prof. S.K. Chatterjee 1939, Abbi 1992, Trivedi 1990 etc. and these works mainly focus on the phonological structure of Echo words. Trivedi 1990 has given eight sets that different Indian languages follow in order to form echo words, but Abbi 2018 in her most recent work has given five strategies which analyze the echo word formation in these languages more precisely. Hence five strategies given by Abbi have been followed to describe the echo words of Bangla in the present paper. Five different strategies that Indian languages undergo to form echo words are enumerated below:

First strategy: Replacing the initial phoneme (mostly consonant) of the base word by a specific phoneme that is unique to the particular language.

Second strategy: Initial syllable of the base word gets replaced by an entirely distinct syllable and the remaining part is canonically copied in the formation of the echo word.

Third strategy: Base word is preceded by its echoed counterpart instead of getting followed.

Fourth strategy: Nucleus of the initial syllable of base word is altered by another vowel in order to form the echo word.

Fifth strategy: Forming echo words by expressive morphology.

2.1 Structure of echo-words in Bangla

Bangla follows the first, third and fourth strategies in order to form echo words. First strategy is the most common device to form echo words and Bangla has few words which undergo the third strategy. The fifth strategy is also followed but it is mainly used to form onomatopoeic words in Bangla.

The first strategy can be schematized by the formula: CVX >CVX-C'VX where CVX and C'VX are base word and its echoed counterpart respectively. C is the initial consonant whose alternation with a distinct phoneme i.e. C' triggers the formation of echo words. VX is the remaining part of the base word which is copied in the echoed part and V represents a pure vowel or a diphthong. It has been mentioned that each Indian language has a more or less constant replacing phoneme, Bangla is no exception. Unvoiced retroflex consonant i.e. /ʈ/ is the most commone replacing phoneme in Bangla echo-formation.

(2) a. dzama-ʈama
          dress-ER   "dress and the like"

     b. pen-ʈen
         Pen-ER   "pen and the like"

     c. tsʰele-ʈele
         boy-ER   "boy and the like"

In Bangla the replacing phoneme is /ʈ/. But this language exhibits some instances in which phonemes also occur as replacing phonemes in echo-formation. Bangla unvoiced bilabial aspirated stop i.e. /pʰ/ occurs as replacing phoneme when the base word has /ʈ/ as its initial phoneme.

(3) ʈaka-pʰaka
     Money-ER"Money and the like/etc."

/pʰ/ is also used to form echo words in Bangla when the echo words tend to express the negligence of the speaker as in (3).

(4) amigan-pʰan korbo na
     I   song-ER do-FUT-1P not
     "I shall not sing songs and the like."

Echo word in example (4) expresses the sense of unwillingness in the part of the speaker in activities like singing and the like.

Bangla echo words formed by following the second strategy is very few in number. In case of such echo formation the initial consonant gets dropped in the echoed part and this echoed part precedes the base word. This can be formulated as CVX > VX-CVX, where word initial consonant i.e. C gets dropped and VX, the remaining part, is copied and precedes the base word.

(5) altu-pʰaltu
     ER -nonsense/rubbish
     "nonsense / rubbish and the like"

Here C= /pʰ/, VX= altu, remaining part of the base word pʰaltu "nonsense/rubbish".

Fourth strategy can be represented by- CVX > CVX-CV'X where CV is the first syllable of the base word CVX, V is the nucleus of this syllable and X is the remaining part which gets copied in the echoed part, V' is the replacer vowel/phoneme or nucleus.

(6) gol - gal
     round-ER   "round shaped and the like"

V=/o/, V'= /a/ so, in (6) nucleus /o/ is replaced by a new nucleus i.e. /a/.

(7) dʰaka-dʰuka
    cover-ER   "cover and the like"

In example (7), V= /a/ is replaced by V'= /u/ to form echo word.

Based on the discussion so far, the intermidiary conclusion could be: (a) if the base word has a vowel as its initial phoneme then in its echoed part a phoneme which is mostly a consonant is added to form echo words and (b) if the base word has a consonantal conjunct in its initial position then the first consonant of that conjunct is replaced and the second consonant is dropped in the echo word.

Rule [a] can be formulated as:

VX > VX-CVX, V is the initial vowel of the base word and a consonant i.e. C is added to form the echo word.

(8) am - ʈam
    mango-ER   "mango etc."

Rule [b] is schematized by the formula:
$C_1C_2VX > C_1C_2VX-C'VX$

(9) prem-ʈrem
    love-ER   "love and all"

3 Types of Bangla echo-words

Echo words have been defined as well as analyzed mainly on the phonological basis. It has been mentioned that echo words express the meaning of "and the like or etc." In other words they echo the meaning of the base words along with the phonological shape of the same. These echo words never occur independently in the language. But Bangla vocabulary contains some echo words which copy or echo the sense of its base word only, that is to say in case of such instances the word that echoes the meaning of base words has its own meaning and independent occurrence but whenever it gets attached to the base word, it loses its original independent status and starts echoing the sense of its base word.

(10) a. dziniʃpɔtro'   "things etc",

    b. 'malpɔtro'   "luggage etc."

    c. 'boipɔtro'   "books and all",

    d 'tsiʈʰipɔtro'   "letters etc."

Bangla word 'pɔtro' originally means "letter", but in 'dziniʃpɔtro' "things", 'malpɔtro' "luggage", 'boipɔtro' "books" it loses its original meaning and echoes the sense of the words it gets attached to. In case of Bangla word 'tsiʈʰipɔtro', 'tsiʈʰi' and 'pɔtro' are synonyms and together they convey the sense of "letters".

Bangla vocabulary contains some pair of words which seem to be echo words in which the initial syllable of the base word get replaced (second strategy of echo word formation seems to get followed here). But if the diachronic study is taken into consideration, it is found that the 'echoed part' of the base word used to be a free morpheme with a lexical meaning but in the course of time it has lost its independent status and at the synchronic level it is serving as a 'echo' to a base word. It may precede or follow the base word.

(11) a. kapor-tsopor   (tsopor< tsupri "small bucket") " clothes etc."

    b. aʃ-paʃ (<Sanskrit arʃe-parʃe, arʃe has no independent occurrence now) "surroundings"

    c. radz-radzra
        "king and other aristocrats"

[radzrais no longer meaningful and serves as an echo to the base word 'radz' meaning "king" though it does not phonetically echo its base word.]

4 Morpho-syntactic Function of echo words in Bangla

Rabindranath Tagore (1909), Prof. S. K. Chatterjee (1939) etc. have mainly studied phonological structures of Bangla Echo words and not focused on its morpho-syntactic function. Lidz (2000) has analyzed Echo word formation in Kannada and mainly emphasized its morpho-

syntactic status in this language. It has been described in his paper that how echo reduplication in spite of being a morphological rule occurs at both lexical and post-syntactic level in Kannada.

(12) a. bagli-annumutʃ-gitʃ-id-e anta hiɭa-biɻa
door-ACC [close-ER]-PST-1P that say-PROH
"Don't say that I closed the door or did related activities."

b. bagli-annumutʃ- id-e-gitʃide anta hiɭa-biɻa
Door-ACCclose-PST-1P-ERthat say-PROH
"Don't say that I closed the door or did related activities."

In (12a) echo reduplication occurs inside of the inflectional elements-'-id' (past tense marker) and '-e' (first personal marker) while in (12b) it occurs outside of the inflectional markers. So, echo reduplication is occurring both at the lexical and phrasal or syntactic level in Kannada respectively. Similarly echo reduplication can be allowed by Bangla words of different grammatical categories post syntactically as well as lexically i.e. this word formation process applies equally to words, subparts of words and entire phrasal categories. Therefore , morpho-syntactic function of Bangla echo words has been analyzed in this paper by following Lidz's work on Kannada Echo words.

4.1 Echo formation of Different Grammatical Categories and Their functions

Bangla words belonging to different grammatical categories undergo echo word formation, though this morphological rule cannot affect Bangla pronouns and adverbs. Nouns, adjectives, verbs, indeclinable words of Bangla can be echoed but they have some restrictions i.e. words belonging to these grammatical categories do not occur with their echoed counterpart in all kinds of syntactic constructions. Echo word formation is generally used in negative statements, prohibition, commands, and interrogative sentences. Some instances of echo formation in declarative sentences are also found.

Echo word formation of Bangla nouns and verbs with different inflectional markers are mainly discussed in the present paper.

4.1.1 Nouns

Echo reduplication applies to both Bangla common nouns and proper nouns, though the instances of echo formation of proper names are rare. Bangla nouns with inflectional markers (case markers/ plural markers) generally allow the echo reduplication inside the inflectional markers i.e. the base is reduplicated first, and then the inflectional markers get added to the word. In such cases echo reduplication occurs at the lexical level.

(13) boi-ʈoigulo gutʃʰijerakʰo
[Book-ER]PL arrange-NF keep-2P-IMP
"Arrange the books etc..."

In (13) Bangla noun 'boi' meaning "book" is reduplicated first resulting in 'boi-ʈoi' "books etc." ,then the plural marker '-gulo' gets added to this reduplicated base and becomes [boi-ʈoi] gulo. So, echo reduplication can be allowed inside the Bangla plural marker.

(14) ami ekʰane kono lok-ʈokke tsinina
I here any [man-ER]ACC know-1p-PRES no
"I don't know anyone here."

Example (14) shows that the Bangla accusative marker '-ke' is added to a reduplicated base i.e. [lok-ʈok]. So, echo reduplication is applied inside the Bangla case marker like plural one.

Echo reduplication cannot be applied outside these above mentioned inflectional markers as *boigulo-ʈoigulo and *lokke-ʈokke are considered as ill formed constructions in Bangla.

Application of echo reduplication to the Bangla proper nouns are found in some specific

constructions such as interrogative and negative constructions in order to express speaker's doubts about some people or their activities, for example:

(15) a. ram-ʈam keu   eʃetsʰe   ki?
         Ram-ER anyone come-3SPRES.PRF what
         "Has Ram or anyone else come?"

    b. adz ram-ʈam keu aʃeni bodʰhoj
        Today   Ram-ER   anyone   come-
        3p.PRES.PRF.NEG may be
        "May be Ram or anyone else has not come today."

4.1.2 Verbs

The echo words of Bangla verbs copy the tense, person markers of the base words.

(16) ami   kʰabo- ʈabo   na
     I   eat-1p.FUT-ER   not
     "I shall not eat and the like."

In (4.1.2.a) echo word of the finite verb 'kʰabo'

(root =kʰa) "shall eat" is '–ʈabo' which has the same tense and person marker or inflectional suffix '-bo', hence it can be said that echo word formation is applied outside these inflectional suffixes or at the post syntactic level.
Echo formation of Bangla finite verbs is mainly found in negative sentences while non-finite verbs with echo formation can occur in affirmative sentences.

(17) ami kʰeje-ʈe jeskule dzabo
     I   eat-NF-ER school to go-1p.FUT
     "I shall go to school after having some food."

Like Bangla nouns, some Bangla verbs allow echo word formation inside their inflectional markers (i.e. at lexical level) and some verbs allow it outside the markers (i.e. at post syntactic level).

(18) ami gaite-ʈaite   parina
     I   [sing-INF]-ER   cannot
     "I cannot sing etc."

In (18) Bangla infinitive marker '-te' is added to the base first then the echo formation is applied to the whole word 'gaite' meaning "to sing". Here, echo word formation is applied outside the inflectional marker or at post syntactic level. Had it been applied to the base first, then it would result in an ill formed construction i.e. *[gai-ʈai]te.

(19) ami kal   kek banije-ʈanijetsʰilam"
     I   yesterday cake [make-ER]PERF.1p.PST
     "I made a cake yesterday."

In (19) echo formation is occurring at the lexical level as at first the verb is reduplicated and then the inflectional markers are getting added to that reduplicated base. So, in case of [banije-ER]tsʰilam, echo formation is affecting the sub parts of the word banijetsʰilam "made".

5 Conclusions

Echo words echo the meaning of the word which they get attached to. They have the power of adding a meaning of 'and the like'. 'et cetera', 'such and such' and also the meaning of 'ignorance or negligence' to the base word. So, they cannot be considered as meaningless and they are not 'empty morphemes'. These echo words rather 'echo morphemes' may be considered as one kind of 'bound base' existing in Indian languages.

Abbreviations

ACC	- Accusative case
ER	-Echo Reduplication
FUT	-Future tense
IMP	- Imperative
INF	- Infinitive
NF	-Non finite
PL	- Plural
PRES	- Present tense
PRES.PRF	-Present Perfect
PERF	-Perfective particle
PROH	-Prohibitive
PST	-Past tense
1p	- First person
2p	- Second person

References

Abbi, A. 1992. *Reduplication In South Asian Languages-An Areal, Typological and Historical study*. New Delhi: Allied Publishers.

Abbi, A. 2018. Echo formations and Expressives in South Asian languages: A probe into significant areal phenomena.*Non-Proto Typical Reduplication:* ed. By AinaUrdze: StudiaTypologica 22. Berlin: Mouton de Guyeter.1-33.

Chatterjee, S.K. 1939. *Bhashaprakash Bangala Vyakaron.* Calcutta: Rupa Publishers.

Emeneau, M.B. 1956. India as a Linguistic Area. *Language,* 32(1), 3-16. http://www.jstor.org/stable/410649 (10th November, 2016).

Lidz, J. 2000. Echo Reduplication in Kannada: Implications for a Theory of Word Formation. *Working Papers in Linguistics*, 6. University of Pennsylvania. 145-166.

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

Tagore, R. 1909. *Shabda-tattva*. Calcutta: Visvabharati.

Trivedi, G.M. (1990). Echo Formation. *Linguistics Traits Across Language Boundaries*: ed. By S.Krishan. Calcutta: Anthropological Survey of India. Ministry of Human Resource Development. 51-79.

CASE MARKING IN NUBRI

Cathryn Donohue

This paper introduces the case marking system in Nubri, using data from both Samagaun and Prok villages. Nubri, a Tibetan language, has ergative and dative morphological cases, but appears to have a dispreference for using morphological case where possible. This paper explores this issue, comparing the variations between the two dialects.

Keywords: case marking, Nubri, ergativity, morphological case

1 The Nubri language

The Nubri Valley is located in the upper Gorkha district of the Gandaki zone in the high Himalayas at the foot of Mount Manaslu in northern-central Nepal. Settled by Tibetans some 400 years ago, this *beyul*, or 'hidden valley' is home to the *Nubripa*, or Nubri people. Most Nubris speak the Nubri language, though in the Kutang area, Kuke is spoken, and in the more recently settled Samdo in the northwest, the villagers speak a language closely related to Kyirung Tibetan. There are also a number of people, who have moved to the Nubri Valley, and whose mother tongue may be altogether different (e.g. Gurung, Manange, etc.). While Tibetan remains the liturgical language and the language of traditional festivals, younger generations are increasingly using Nepali, even between themselves, as it is the language of the screen and contemporary songs, and a language symbolic of economic opportunity and modernity. Further, government health and education assistants assigned to the area typically do not speak Nubri resulting in widening domains of language attrition.

There are approximately 2000 people across the Nubri Valley (Simons &Fenig 2018), with 800-1000 of those located in Samagaun, the largest of the Nubri villages. There are reportedly 500 monolingual speakers of Nubri in the valley, though it is unclear how many of the 2000Nubri people speak primarily Nubri, and it is yet to determine the dialectical variations of the Nubri language systematically.Ethnologue reports four main dialects (Sama, Lho, Namrung and Prok), though it is universally accepted in the Nubri

valley that the dialect spoken in Samagaun is the most distinct from other varieties, which our fieldwork has confirmed. With the exception of a couple of short word lists, and a recently published lexicon (Dhakal 2018), Nubri remains undocumented.

The data presented here are primarily from the variety of Nubri spoken in Sama village and are based on fieldwork carried out by the author over six trips to Nepal during 2016-2018. The Nubri project started in 2016 together with Mark Donohue, and the data here builds on the foundational work carried out jointly.

2 Case systems

Our work on Nubri started out following typical elicitation methodology of translating sentences from English/Nepali/Tibetan into Nubri. In the initial stages, it appeared to be a typically ergative Tibetic language (e.g.Tournadre 2013). However, after collecting naturalistic discourse of varying kinds (conversation, narratives, pear story recounts from several speakers, real-time pear story narratives etc.), it has become clear that case marking is rarely used and that case is not preferred in general.

However, Nubri does have two structural cases: ergative (-*gi/yi*) and dative (-*la*),prototypicallymarking the transitive subject and indirect object. Examples of some prototypical predicates are given below.

(1) a. nga   zei   yin
     1.SG   eat   AUX
     'I ate.'

   b. nga   shau   zei   yin
     1.SG   apple eat   AUX
     'I ate the apple.'

   c. nga-i   shau   zei   yin
     1.SG-ERG apple eat   AUX
     'I ate the apple.'

Both (1b) and (1c) are considered grammatically correct, but the sentence without the ergative case

marking (1b) is what is commonly found in natural discourse.

The dative case is prototypically found on the recipient argument ('indirect object') in a verb such as 'send', illustrated in (2) below.

(2) mo   kuttshap-la   yige tang   so
    3SG.F ambassador-DAT letter send   PRF.
    'She sent a letter to the ambassador.'

Note that it is less commonly found, but not ungrammatical to have the ergative case marking the agent in the sentence in (2) as in (2′) below.

(2′) mo-yi   kuttshap-la   yige tang so
    she-ERG   ambassador-DAT letter send PRF.
    'She sent a letter to the ambassador.'

Related Tibeto-Burman languages have shown a complex of contexts that may affect the presence of the ergative case (e.g. Bumthang, Donohue & Donohue 2016), so we may anticipate that the tense/aspect of the event, the affectedness of the object, or the volitionality of the subject, may influence the presence of the ergative case but this seems not to hold true in SamaNubri as illustrated in (3). In (3a) and (3b) show that changing the aspect of the clause from progressive to perfective does not result in the ergative case marker. (3b) and (3c) show that the telicity of the event neither forces the use of the ergative case and (3d) and (3e) show that forcing a highly volitional subject does not result in the use of the ergative case marker.

(3) a. o   magyen di   chörpi tup nu
      DET woman DET cheese cut PROG
      'The woman is cutting the cheese.'

   b. o   magyen chörpi tup   so
      DET   woman cheese cut   PRF
      'The woman cut the cheese.'

   c. mo   chörpi yölu   tup   so
      3SG.F cheese some   cut   PRF
      'She cut some of the cheese.'

   d. o   magyen dzöl-ne   chörpi tup
      DET woman   accidentally cheese cut
      so
      PRF
      'The woman cut the cheese accidentally.'

   e. o   magyen chörpi kang tsu-ne tup so.
      DET woman cheese deliberately cut PRF
      'The woman cut the cheese deliberately.'

The dative case is morphologically identical to the semantic use of the dative, used to mark locative or allative arguments as shown in (4).

(4) a. bõ   sha   so
      girl   go   AUX
      'The girl went.'

   b. bõ   lungpa-la   sha   so
      girl   village- DAT   go   AUX
      'The girl went to the village.'

However, occasionally it is used to mark the 'object' of a transitive verb. It is this particular usage of the dative case that I will focus below.

3 On the variable use of the dative case

While the use of the ergative marker in Sama Nubri is sparse, core objects of a number of transitive verbs bear the dative case. Consider the examples in (5).

(5) a. mo   shel   tup   so.
      3SG.F glass   cut   PRF
      'She cut the glass.'

   b. shel   mo-la   tup   so.
      glass   3SG.F-DAT   cut   PRF
      'The glass cut her.'

In (5a) we see that there is no case marking on either the subject (*mo*) or the object (*shel*), while in (5b), the object is marked with the dative case, *-la*.

It is striking that in (5a), the subject 'she' is a good prototypically agentive animate noun phrase and the object is a good prototypically inanimate noun phrase, while in (5b) the reverse is true.

This is in line with the animacy hierarchy, first proposed by Silverstein (1976) for explaining morphologically split ergativity in Australian languages. The examples in (6) show an extreme opposite of a highly animate subject coupled with an inanimate object in the (a) example, then the reverse in the (b) example. In this instance of 'animacy reversal' we find the object marked by the dative case.

(6) Animacy hierarchy relevant for Nubri

Highest	*First/Second person*
pronouns	
↑	*Third person pronouns*
|	*Nouns with human referents*
↓	*Nouns with animate*
referents	
LowestNouns with inanimate referents	

The case marker can be omitted when the canonical word order, SOV, is adhered to, and there is a clearly suitable 'subject' candidate, such as 'glass' as a cause of the cutting event in (7).

(7) shel   mo   tup   so.
    glass   3SG.F   cut   PRF
    'The glass cut her.'

Once a non-canonical word order is introduced (for discourse reasons), the case marker is used, as shown in (8) below.

(8) a. mo-la   shel   tup   so.
       3SG.F-DAT   glass   cut   PRF
       'The glass cut her.'

    b. *mo   shel   tup   so.
        3SG.F   glass   cut   PRF
       'The glass cut her.'

The examples in (8) show us that non-canonical order is possible when the object bears the dative case, but not possible without case on the object, as the ungrammaticality of (8b) shows.

However, the relative animacy heirarchy of the participants does not always change the case marking in transitive verbs. In the examples in (9–11) we see that the objects of 'hit' always bear the dative case, as the (b) examples show.

(9) a. mo   yak-la   zhü   so.
       3SG.F   yak-DAT   hit   PRF
       'She hit the yak.'

    b. *mo   yak   zhü   so.
        3SG.F   yak   hit   PRF

(10) a. mo   kho-la   zhü   so.
        3SG.F   3SG.M-DAT hit   PRF
        'She hit him.'

     b. *mo   kho   zhü   so.
         3SG.F   3SG.M   hit   PRF

(11) a. yak   mo-la   zhü   so.
        yak   3SG.F –DAT   hit   PRF
        'The yak hit her.'

     b. *yak   mo   zhü   so.
         yak   3SG.F   hit   PRF
        'The yak hit her.'

It has been reported that, in classical Tibetan, the dative case may be used to mark objects depending on the nature of the verb: objects of change of state verbs, such as 'kill' or 'cut' do not bear case, while those of verbs involving surface contact (such as 'hit') are marked with *la* (DeLancey 2003:259). It is similarly true in Nubri that the verb 'hit' marks its object with *la*, although the verbs with the semantic feature of change of state behave differently.[1]

Verbs of contact have been identified as a group that requires objects to bear case in related languages (e.g. Classical Tibetan, DeLancey 2003:259). This holds true for many verbs in Nubri such as 'hit' as we saw in (9-11), but it is also used across the board for a verb like 'look at' as shown in (12).

(12) a. mo   yak-la tei   so.
        3SG.F   yak-DATlook.at:PRF   PRF
        'She looked at the yak.'

     b. mokho-la   tei   so.
        3SG.F 3SG.M-DAT look.at:PRF   PRF
        'She looked at him.'

     c. yak   mo-la   tei   so.
        yak   3SG.F-DAT look.at:PRF   PRF
        'The yak looked at her.'

One type of verb such as 'see' shows some variation in the case marking consistently related to the (relative) animacy of the arguments as shown by the absence of a case marker in (13a) and the presence of a case marker in (13b).

(13) a. mo   khi   tung   so.
        3SG.F   dog   see   PRF
        'She saw the dog.'

     b. khi   mo-la   tung   so.
        dog   3SG.F-DAT   see   PRF
        'The dog saw her.'

The whole paradigm of interactions of different animate subjects and objects as arguments of the

verb 'see' is given in Table 1, which assumes clauses with canonical SOV word order. Non-canonical word order requires case marking of the object. With a verb like 'see' there can be no inanimate subjects, hence the missing row.

Table 1: Object marking with 'see' in Sama Nubri

OBJ \ SUBJ	Local person	3rd person	Human	Animate	Inanimate
Local person	la	(la)	–	–	–
3rd person	la	la	(la)	–	–
Human	la	la	la	(la)	–
Animate	la	la	la	la	–
Inanimate	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*

The selective use of the object marker with 'see' is clearly determined by properties of the arguments participating in the event. There are several interesting points to note from Table 1:

i. If the animacy of the subject is lower than that of the object, the object must bear case;

ii. If the animacy of the subject and the object are equal, then the object bears case;

iii. Objects in clauses with higher animate subjects do not typically bear case but may optionally bear case if the object is just one 'step' less animate than the subject.

4 Dialectal variation of object marking

The Nubri language is quite different at the other end of the valley in many ways, including case marking patterns. Focusing on case marking for now, in the Prok variety of Nubri, as in Sama, we have verbs such as 'hit' and 'look at' that uniformly mark objects with the dative case, but we do also find variations with verbs like 'see' as shown in Table 2, again assuming canonical word order. While similar, this is crucially different from SamaNubri. What we see here is a general non-preference for non-human referents to be marked with case. Further, even in equal animacy contexts non-pronominal referents typically do not bear case except in the case of animacy

reversal, then human objects bear case. Moreover, while case is required on local objects in clauses with third person pronominal subjects (animacy reversal), case is optionally allowed on all other pronominal objects.

Table 2: Object marking with 'see' in Prok Nubri

OBJ \ SUBJ	Local person	3rd person	Human	Animate	Inanimate
Local person	(la)	(la)	–	–	–
3rd person	la	(la)	–	–	–
Human	la	la	–	–	–
Animate	la	la	la	–	–
Inanimate	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*

5 Concluding remarks

This paper is the first to report on case marking in Nubri. With a focus on Sama Nubri, I show that the use of case is not just sparse, but interestingly controlled in its variation which crucially relies on the type of verb, word order, and the relative animacy hierarchy of the two arguments. I also show that the Prok variety of Nubri is quite different to Sama Nubri.

Such variability in case marking data provide challenges to theories of case which assume that case marking is fixed and assigned on purely structural facts. These data suggest that minimally the animacy of the noun phrase arguments must be taken into consideration for at least a subset of the verbs, in order to correctly understand how case marking is assigned in Nubri.

References

DeLancey, Scott. 2003. Classical Tibetan. in G. Thurgood and R. LaPolla (eds), *The Sino-Tibetan Languages*. London: Routledge. pp. 255-269

Dhakal, Dubi Nanda. 2018. *A Nubri Lexicon*. Munich: Lincom Europa.

Donohue, Cathryn, and Mark Donohue. 2016. Onergativity in Bumthang. *Language* 92.1: 179-188.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In R.M.W.Dixon (ed.) *Grammatical Categories in Australian Languages*, Canberra: Australian Institute of Aboriginal Studies, pp. 112-171.

Simons, Gary F. and Charles D. Fennig (eds). 2018. *Languages of the world, 21st edition*. Dallas, TX: SIL International. Online version http://www.ethnologue.com

Tournadre, Nicolas. 2013. The Tibetic languages and their classification. In T. Owen-Smith and N. Hill (eds), *Trans-Himalayan Linguistics*. Mouton: De Gruyter. pp. 105-130.

---

[1] Note that the verb for 'hit' *zhü* requires its subject to be volitional, so it is not possible to check a sentence where the subject and object are both inanimate. Instead a different verb is used which requires its subject to be a non-volitional causer of the event.

LANGUAGE SHIFT IN NEWAR: A CASE STUDY IN THE KATHMANDU VALLEY

Bhim Lal Gautam

This paper explores the patterns of language shift in Newar, the ethnic indigenous language community living in the Kathmandu Valley. The research focuses on language contact situations in different domains viz. social, cultural, personal, and official as well as media related activities where the informants are asked to use different languages along with the use of their own mother tongue i.e. Newar. This socio-ethnographic research aims at providing some clues as to how the discovery of a minority language triggers changes in representations and attitudes.

Keywords: language shift, causes and impacts, ideology, globalization

1 Introduction

The Newars are commonly believed to be the indigenous inhabitants of the Kathmandu Valley, and are recognized as such in the historical descriptions of the Nepal Valley (Malla 2015). The civilization and culture of the Kathmandu Valley are identified with the Newar civilization and culture. Nepali (1965) has observed that "the Newars are a people with a high degree of material culture and a distinctive social organization. "The origin of the Newars, however, still remains uncertain and the proto identity of the Newar people continues to be disputed among various schools of thought. Grierson (1909) in the *Linguistic Survey of India* had proposed the popular hypothesis that there were two branches of migration along the Himalayas from east to west, while Chatterjee (1950) and Regmi (1960) assign the first branch to North Assam (the Newars included) and the second branch to Outer Mongolia and Tibet in the north. Scholars thus have not been able to connect the Newars with the migration pattern proposed in the *Linguistic Survey of India.*

The Newar speakers are concentrated in the three major cities of the Kathmandu Valley viz. Kathmandu, Lalitpur and Bhaktapur and well-defined range of urban settlements across the country. The CBS Census Report of 2001 gives a total of 12,45,232 (5.4%) ethnic Newars, and 8,25,458 (3.03%) mother-tongue speakers which indicate a decline of 33.7% inactive speakers.

Census 2011 reports there are 13,21,933 ethnic Newars and 846557 (3.2%) mother tongue speakers. Some view this trend as alarming, but the Newars continue to use their language extensively in many domains of socio-cultural contexts, trade and commerce, education, literature and massmedia. The Newars are also a highly literate community.

Newar also known as Nepal Bhasa, is a Sino-Tibetan language spoken by the Newar people, the indigenous inhabitants of Nepal Mandala, which consists of the the Kathmandu Valley and surrounding regions in Nepal. Although "Nepal Bhasa" literally means "Nepalese language", the language is not the same as Nepali , the country's current official language. The two languages belong to different language families (Sino-Tibetan and Indo-European respectively), but centuries of contact have resulted in a significant body of shared vocabulary. Both languages have official status in Kathmandu Metropolitan City. Newar was Nepal's administrative language from the 14th to the late 18th century. From the early 20th century until democratization, Newar suffered from official suppression. From 1952 to 1991, the percentage of Newar speakers in the Kathmandu Valley dropped from 75% to 44%, and today Newar culture and language are under threat. The language has been listed as being "definitely endangered" by UNESCO. The typological and genetic classification of Newar/Nepal Bhasa has been controversial for several reasons. However, contributing factor has to do with the long period of contact with Sanskrit, Prakrit and other Indic languages resulting in considerable lexical and grammatical borrowings.

Although Newar population has increased between the 1952/54 to 2011 censuses, the Newar group no longer maintains numerically the highest position in Kathmandu district in the 2011 census. It is simply because a large number of Hill Brahman and other ethnic groups migrated into the Kathmandu Valley recently because of the Maoist insurgency in Nepal from 1996 to 2006 for

reasons of security and employment.

Table 1: Newar mother tongue population in different census

SN	Census Year	Total Population	Percentage in Total population
1	1952/54	383184	4.65%
2	1961	377721	4.01%
3	1971	454979	3.94%
4	1981	448746	2.99%
5	1991	690007	3.73%
6	2001	825458	3.63%
7	2011	846557	3.20%

In addition, every year a lot of people migrate to the Kathmandu valley searching for jobs, education, business etc. and eventually settle in the valley.

Table 2: Newar population in two censuses

SN	Districts	2001			2011		
		Total	Male	Female	Total	Male	Female
1	Kathmandu	295439	146279	149160	383136	190297	192839
2	Lalitpur	138938	68386	70552	155604	76703	78901
3	Bhakatapur	126592	63098	63494	138873	69239	69634
Total		560969	277763	283206	677613	336239	341374

Table 2 shows that Newar population in the Kathmandu valley is increasing but if we compare the ratio of total population, we can easily guess that it is decreasing day by day because of development, communication and urbanization.

2 Methodology

2.1 Theoretical framework

This research was mainly based on the data collected through questionnaires which were developed in 2014 by the researcher in DDL, University of Lyon 2, France and their pilot testing was done in 2015 in Kathmandu. The questionnaires were translated into Nepali and administered to the concerned informants. Moreover, this paper is also based on some existing literature related to language contact and language ideologies while describing and analyzing the data. The data were collected and analyzed following the mixed method approaches i.e., quantitative and qualitative.

2.2 Tools for the study

a. Survey questionnaires

There were 45 indicators in the questionnaires altogether containing metadata information and questions for language use and attitude. The collected data was analyzed by following the recent developments of language contact, ideology and sociolinguistic studies.

b. Focus group discussions (FGD)

There were four FGDs in four different research sites in this research. One FGD was conducted for each of the Linguistic/speech communities involved. There were about 5-7 members in each of FGD. Members were purposively selected from sex, age, occupation and education strata.

c. Interview

At least 8 individual interviews were conducted in order to meet the need and objective of the research focusing on the language attitude and ideology in multilingual context.

2.3 Selection of informants

The data was mainly collected from the people of Newar language communities like housewives, teachers / academicians / monks, politicians / language activists, businessmen/shopkeepers, trekking guides / workers/ vendors, students. All the informants were selected on A1, A2 and A3 group classifying into male / female, literate / illiterate whenever possible. A1 belongs to the age between15-30, A2 belongs to 31-55 and A3 belongs to 56 and above. Figure 1 shows the brief information about the informants.
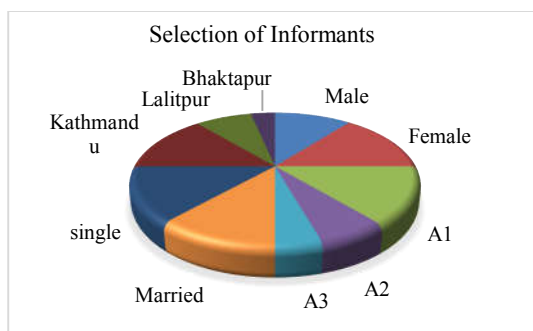
Figure 1: Selection of informants

2.4 Data collection

The primary data was collected with the help of questionnaires and recording FGDs and interviews. Questionnaires were used for language use and attitudes and FGDs and interviews for sociolinguistic analysis of language shift in Newar. The source of data was based on researcher's informal field study like social talking, business talking, debate and conversations etc. rather than written sources. The secondary data were collected from different libraries and sources available.

3 Patterns of language shift

This part mainly deals with the trends of language shift focusing on the diverse domains of language use and attitudes among the Newar language speakers living in the Kathmandu Valley. The researcher administered questionnaires containing 45 indicators in different five selected areas of the Kathmandu Valley including some FGDs and informal interviews. The trends and patterns of shift noticed in Newar languages are described below.

3.1 Informal situations

In informal situations people do various activities informally without being conscious and caring about the outer community. The informal situations in this speech community comprise two types of activities: behavioural and personal. They are briefly discussed as follows:

a. Behavioural activities

Behavioural activities in this research refer to the activities that indicate different psychological behaviours of the informants. They include the activities like making friends, different reading and writing activities, making telephone calls, talking with different people, shopping, passing exams and so on.

Table 3 shows that The Nepali language is used highly in almost all the domains in comparison to Newar, English and Hindi. The influence of English is much higher than that of Hindi because of education, globalization and tourism among the Newar speaking community in the Kathmandu Valley. Newar is highly used for making friends,

shopping, telephone calls and talking with workers rather than other activities. The use of English is very high for reading/writing, passing exams, getting jobs and talking with teachers and academicians.

Table 3: Patterns of language use in behavioural activities (N=45)

Activities	Languages				
	Newar	Nepali	English	Hindi	Others
1. Making friends	73.33%	93.33%	15.55%	6.66%	
2. Shopping	71.11%	97.77%	6.66%	35.55%	
3. Making telephone calls	80%	93.33%	22.22%	6.66%	
4. Talking with workers	60%	93.33%	22.22%	13.33%	
5. Talking with teachers/ professors	22.22%	93.33%	35.55%		
6. Talking with academicians	28.88%	95.55%	26.66%		2.22%
7. Getting a job	13.33%	77.77%	40%		
8. Reading and writing	35.55%	91.11%	46.66%		
9. Passing Exams		66.66%	42.22%		

b. Personal activities

Personal activities in this research mean those activities which are connected to the personal and interpersonal activities of the informants. They include activities like joking, singing, praying, bargaining, abusing, telling stories and so on.

Figure 2: Language use in personal activities

Figure 2 shows that Newar, Nepali and English are used in most of the activities like joking, singing, praying, bargaining with highest frequency than Hindi and other languages. Here the use of Hindi is very low because of these informal personal activities.

As marked in the Figure 2, the Newar language is dominant among the Newar mother tongue speakers in the activities of joking, quarrelling, verbal abusing, family gathering and village/community meetings or gatherings. However, the Nepali language is more domin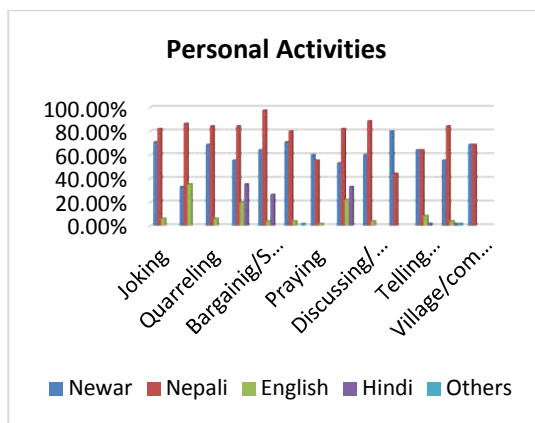ant in the activities of singing inside and outside, counting, discussing and shopping/marketing. Presence of Hindi is clearly noticed in the activities of shopping, singing inside and outside. English is used to in most of the activities except family gathering. The figure also demonstrates that in the activities of village/community meeting, telling stories to children and telling stories to others, both Newar and Nepali are found parallel in use.

From the close observation of the figure above and my interview transcripts (FGD/individual) collected from the field, it is apparent that the Nepali language is gradually being dominant among the Newar mother tongue speakers. As a Newar man (56) commented, "We use the Newar language at home and community since we are all Newars here; but when we go to our children's schools, or go to *Shahar* (Kathmandu city) for shopping, we use Nepali, so I know both languages and need both for different situations." (Interview, February 2018) Though a small segment, the comment of this Newar man indicates how language use is in shifting trend from Newar to Nepali even in the personal activities.

3.2 Formal situations/activities

The table 4 below shows the use of language among Newar mother tongue speakers in the formal situations and activities. As the table indicates, Nepali is found highly dominant among the Newar mother tongue speakers in the formal situations and activities. In the activities of reading and writing, seeking jobs, attending exams, talking to teachers and intellectuals, language use in offices and work places as well as

social political activities, Nepali is most influential.

Table 4: Patterns of language use in personal activities (N=45)

	Situations	Languages				
		Newar	Nepali	English	Hindi	Others
1	Office/ workplace	33.33%	84.44%	15.55%	4.44%	
2	Political/ social gathering	44.44%	88.88%	2.22%		
3	Public activities/ fun fair	51.11%	88.88%	4.44%		
4	Administration	6.66%	97.77%	8.88%		
5	Strangers	8.88%	97.77%	4.445	2.22%	

As commented by many aged literate people of the typical Newar community in Kathmandu, the fundamental reason for the people shifting from Newar to the Nepali language is that it is not a medium of instruction in schools. The table also indicates that the Newar language occupies no space among the native speakers of this language in the administrative works. In some cases, English is found more dominant than the Newar language. The presence of Nepali and English among the Newar speakers in this manner implies the process of shifting from Newar to Nepali and English.

The dominance of the Nepali language among the Newar mother tongue speakers can be observed in the personal stories as well. As a Newar woman (65 years) comments, "Children these days speak either Nepali or English; they use Newar language only when we Newar speaking parents and grandparents talk to them; they do not prefer to speak Newar; they mix up Nepali and English words even if they speak Newar at all" (Interview, February 2018).

3.3 Religious and cultural activities

Table 5 demonstrates the trend of language use among newer mother tongue speakers in religious and cultural activities. Presence of Newar and Nepali can be seen in all the activities listed here under the various domains of religious and cultural activities. However, use of the Newar language is dominant in each and every case. Presence of English is very rare in cultural programs, marriage

ceremony and cultural festivals and the use of Hindi is almost zero in the data.

Table 5: Patterns of language use in personal activities (N=45)

Activities		Languages				
		Newar	Nepali	English	Hindi	Others
1.	Religious Festivals	38	12			
2.	Cultural Programmes	36	22	1		
3.	Death Ceremonies	39	13			
4.	Marriage Ceremonies	39	19	1		
5.	Birth Ceremonies	40	15			
6.	Cultural Festivals	39	22	1		

3.4 Family and friends

Table 6 demonstrates the language use trends among people speaking Newar as mother tongue while communicating with family members and friends.

Table 6: Patterns of language use in personal activities (N=45)

Persons	Languages				
	Always Newar	Newar Less Nepali	Nepali less Newar	Always in Nepali	Nepali and English
Father	87%	8.88%		11.11%	
Mother	84.44%	6.66%		8.88%	
Brother/ Sister	77.77%	2.22%	4.44%	15.55%	
Spouse	53.33%	2.22%		8.88%	
Friends at home	51.11%	26.66%	2.22%	20%	
Friends outside	15.55%	24.44%	15.55%	31.11%	13.33%
Neighbour at home	55.55%	13.33%	8.88%	22.22%	
Neighbours outside	31.11%	26.66%	11.11%	31.11%	

We see the presence of Nepali and Newar in all domains. While communicating with friends outside home, use of Nepali is more dominant; but Newar is more dominant in the case of communication with friends at home. Talking with the immediate relatives of family (spouse, parents and siblings), Newar is dominant.

3.5 Media and entertainments

Table 7 demonstrates the trend of language use among Newar mother tongue speakers in the activities under the domain of media and entertainment.

Table 7: Patterns of language use in personal activities (N=45)

	Activities	Languages				
		Newar	Nepali	English	Hindi	Others
1	Watching Movie/ Serial	16	40	15	34	9
2	Watching News	10	44	8	13	
3	Listening Music	26	38	17	35	
4	Listening Radio/News	19	44	11	5	
5	Listening Interview	18	44	10	10	
6	Reading Newspaper	13	41	15	2	
7	Reading horoscope	9	42	15	1	

Table 7 shows that Nepali is dominant in media and entertainment activities compared to Newar itself. As the data indicate, Newar mother tongue speakers use the native language in the activities of watching TV serials, TV news and listening to music; however, it is less dominant in these activities. In the activities of listening to news, listening and watching interviews, reading newspapers and reading horoscopes, Nepali retains its wide dominance among the Newar mother tongue speakers. Hindi is found to occupy more space in the activity of watching TV serials and listening to music. English marks presence in media and entertainment activities including watching TV serials and reading newspapers. However, these languages (Hindi and English) are not so wide as Nepali is in media and entertainment activities among the Newar mother tongue speakers.

4 Language contact and intergenerational shift

Multilingualism, language contact and language shift are the inherent miracles in Nepal. Language shift, sometimes discussed as language transfer or language replacement or assimilation, is the process whereby a speech community of a language shifts to speak another language, usually over an extended period of time. The language shift may have different effect on language community. There may be cultural shift along

with language shift; and some different language can emerge. Language shift is one of the effects of globalization. Nepal's labour migration and market force have their effect on language shift. Nepal's labours working in different countries are acquainted with the languages used in those countries. As a result, there is language shift. There may occur communication problem as well among the speakers of the same language if they are of different ages and reside in different places.

Most of the adult people of Newar community interviewed during the fieldwork said that they use Newar language while talking to their older generations, but Nepali while communicating with the younger ones. And, such a shift found among younger generations towards Nepali and even English (to some extent) has been reinforced along with the growing influence of modern media and technology. The generation gap among the different age groups can be found in a remark made by a Newar man (56) who said, "I myself and my wife like Newar TV programs, but my daughter prefers Hindi serials, she always watches these serials, and for son enjoys sports mostly in English language" (Interview, February 2018).

However, in some cases, the younger generation people (15-25 years) were found using the mother tongue in all the personal activities covered by this study. Newar speakers have used their native tongue in the domains of cultural, religious and some of the formal situations as well. However, data indicates that the younger generation of mother tongue speakers are gradually motivated to other 'dominant' languages under the influence of current globalization, education, migration, business and communication, and technologies and the media.

5 Causes and impact of language shift

This section explores the causes and impacts of such shifts in languages spoken by Newar people in the Kathmandu Valley. The analysis of FGDs, individual interviews and field observations demonstrate the following causes and impacts.

5.1 Media, migration and marriage (M3)

The analysis of the data demonstrates that there are many causes that promote language contact. Media, migration and marriage (M3) are found to retain some vital impact in contact and shift. Growing consumption of media, no matter whether that be electronic or print, makes the local mother tongue speakers get exposed to regional and global languages. As it is observed in the Nepal's linguistic landscape, the natural influence of Nepali, Hindi and English medium TV channels is apparent. Such influence could be observed in people's growing consumption of Nepali channels especially for news, entertainment and information of the diverse social, cultural and political aspects of the state; Hindi channels especially for entertainment; and English channels in order to get entertainment, sports and international/global exposure. On the other hand the growing use of internet on various social networking sites like Facebook, Twitter, Skype, Viber, etc. have forced people to shift towards English and many other foreign languages through smart phones, pads and computers.

Migration, as pointed out above, is another vital cause of language contact and shift. In the Kathmandu valley, the city of Kathmandu has almost 42 per cent internal migrants from both rural and urban areas of other districts (CBS 2011). The absent population of Nepal has been a major issue in demographic, social and economic aspects of the country. The absent population reported in 2011 was 1,921,494, a big jump from the number of 762,181 of the census of 2001.Nepal currently results widespread internal as well as external migration rates in the history. As my analysis points out, internal migration has provided the space for promoting language shift from local languages to Nepali, English and Hindi languages because of the heavy migration of different people in the Kathmandu valley.

Marriage is another cause for migration that often promotes language shift. Marriage that occurs between the members of inter-lingual background is quite imperative to understand how it promotes language shift. Sony, a Newar speaking woman's experience often demonstrates that how her marriage to a Nepali speaking Tamang man ultimately pushed her to shift from Newar to Nepali. Thus, the data indicate that marriage has become an important cause of language shift

mostly in the case of female participants of diverse ethnic groups.

5.2 Education

The next indicative cause that encourages language shift is education/educational access. The state's ideology to expose the young citizens with more dominant national, regional and international languages promotes the language shift especially on the young citizens. The analysis indicates that Newar explored in this study is extensively used in its respective local contexts. However, the language has not been (well) recognized in formal schooling and academic institutions. Even some practices of educating young children in their mother tongues have been appropriately materialized due to the lack of public interest, concrete policies on mother tongue education, and parental desire to expose their children with more dominant languages such as Nepali and English. Since "English-prioritized schooling access for the children is often perceived as the symbol of better future, better social status and economic soundness of the household" (Devkota, 2018, p. 111), parents are more motivated to make their children to learn English and Nepali languages instead of their mother tongues.

In Newar dominant community of the Kathmandu Valley, many of my participants often explained how their children have been extensively exposed to Nepali and English languages in place of their mother tongues. Badri Lal Maharjan (75 yrs.) from a typical Newar speaking community said, "In the very past, there was the provision of the Newar language subject in school curricula along with Nepali and English, but now schools don't teach that, our children do not study the Newar language anywhere formally, they learn Nepali and even more English instead, so how they could learn the Newar language" (Interview, February, 2018).

In FGDs, that took place in three major areas of the Kathmandu Valley, the participants often explained how their children have been extensively exposed to Nepali and English in place of their mother tongues.

5.3 Travel and tourism

Nepal is not only linguistically and culturally diverse, but also ecologically and geographically distinct. Due to its diversity in geography as well as rich and unique climate, Nepal has a wide range of plant and animal species, famous for its flora and fauna, beautiful rivers and lakes. Nepal is the land of world's highest peak, *Sagarmatha* (Mt. Everest), *Lumbini*, the birth place of Buddha, *Janakpur*, birthplace of Janaki and *Pashupatinath* temple where visitors from across the world visit for different purposes. Every year thousands of people from around the world visit Nepal for many reasons. Some of them are on holiday, while some of them want to study, the languages and cultures of the people and the country's natural resources. For all of these visitors, whatever their national languages, English plays a significant role as a medium of communication, from booking tickets and hotels to arranging travel and trekking during their stay in Nepal.

All stakeholders associated with the travel and tourism industry, such as hotel, restaurant and trekking entrepreneurs, must have knowledge of English in order to communicate with the tourists. Not all people involved in this business are highly educated. Some are illiterate and run small lodges and restaurants on their respective areas. They automatically develop the competence of foreign languages like English, Hindi, Chinese, etc. through the direct contact with foreigners. Thus the travel and tourism is one of the important factors of language contact and shift from local language to many foreign languages. Many Newar people involved in this field understand and speak these languages.

5.4 Market force and economic benefits

Market force is another reason that is found intricately connected to the economic benefits of the individuals. This research demonstrates a number of examples from diverse study sites where people are using and shifting their language because of business and economic purpose. A businessman (Ravi,47 yrs.) at Patan said, "I like to speak my own language but I have to speak Nepali, English or Hindi otherwise people do not come to my shop" (Personal Interview, 2018: my translation). The story of Ravi can be understood

how language shift is guided with the market force particularly business.

Current market force is seen to be imperative in language shift in different ways. Growing trend of external migration of Nepali citizens in the international labour markets encourages them to learn a particular language(s) that benefits them in a particular social context; no matter it is extensively for instrumental purpose in the beginning. It ultimately ensures language shift. Similarly, another Newar speaking man (38) from Bhaktapur commented, "Where are our languages, I mean Newar, Tamang, Gurung.…in media we listen to Nepali and English, even Hindi rather than our local languages, we read English, Nepali, Hindi …in the advertisements and manuals of materials we buy at home…we have to go to other countries for work or study, so what happens if we don't learn these languages?" The participants' comments as such clearly hint why people get motivated to the languages such as Nepali in the national context and English or other dominant languages under the influence of current market force.

5.5 Political-ideological intervention

Political-ideological intervention of the state is another powerful cause in promoting language shift through contact. The 1959 and 1962 constitution of Nepal confers the status of the national language to Nepali. During this period, Nepali has taken great advances to raise itself to the status of the national language. Although studies on the comprehension and the use of Nepali by non-Nepali speakers are far and few between completefeasibility seems to have driven more and more non –Nepali speakers to use and understand it in their day to day transactions, inter-ethnic communications and above all in their communication with the channels of the local and national administration. Since the very instigation of modern Nepal, the Nepali language was promoted as the language of unity, the language of social harmony and national integration. As well, the Nepali language was also intensively employed in schooling/education, media and formal communication in the state. Such intervention has ideally promoted the Nepali language throughout the nation.

Due to the lack of a concrete plan by the Nepalese government regarding the development of the ethnic languages, the English language, along with Nepali, has been predominant in school curricula, both in the rural and urban areas of Nepal. The learning of English provides Nepalese with opportunities to obtain jobs in various national and international governmental organizations and in the media. Therefore, a large section of the Nepalese people is attracted to the English language more than other local languages.

As it is explored from the interviews and FGDs, the growing shift to Nepali and English language is because of its extensive use and applicability in the formal situations, media, education and other formal fields in the nation. The narratives of the respondents belonging to diverse the socio-economic backgrounds demonstrate that their shift to Nepali and English is mostly related to fulfilling the pragmatic purposes.

6 Summary

The data presented in this study shows that mother tongue is highly used in cultural and religious activities. Nepali is dominantly used in social, official, ceremonial and media related activities. English and Hindi languages are used in media, ceremonial and official activities. The influence of English is much higher than Hindi among Newar people which indicates the influence of globalization and western traditions. Newars have been directly involved in official and academic activities as well as tourism which motivate them to contact with many foreigners in English and Nepali rather than other languages. A shift in a language often brings about a shift in identity and there may be resistance to adopting a new language. The new language and the new identity may be actively promoted or persuaded. Newars living in the capital city have been influenced directly and indirectly by the globalization and international linkage and communication. Moreover, they have been involved in various social, cultural and ceremonial activities with the new mixed society which motivates them to shift into new target languages from the ancestral source language. In this context, this study is connected with the socio-political factors/variables where different

language communities/speakers share different contexts and situations. So the multilingualism in the Kathmandu Valley has become an obligatory part of people living in this city. Existing political, social and economic factors contribute to language use and attitude. Nepali being the dominant language in the capital city, the lingua franca of the country and English being the international language of various purposes is becoming more valuable and influencing in Newar speaking community warn language shift and endangerment.

References

Central Bureau of Statistics. 1952/54. Population Census, National Planning Commission (NPC), Kathmandu, Nepal.

Central Bureau of Statistics. 1961. Population Census, National Planning Commission (NPC), Kathmandu, Nepal.

Central Bureau of Statistics. 1971. Population Census, National Planning Commission (NPC), Kathmandu, Nepal.

Central Bureau of Statistics.1981. Population Census, National Planning Commission (NPC), Kathmandu, Nepal.

Central Bureau of Statistics. 1991. Population Census, National Planning Commission (NPC), Kathmandu, Nepal.

Central Bureau of Statistics. 2001. Population Census, National Planning Commission (NPC), Kathmandu, Nepal.

Central Bureau of Statistics .2011. Population Census, National Planning Commission (NPC), Kathmandu, Nepal.

Chatterjee, Suniti K. 1950. *Kirata-Jana-Kriti : The Indo-Mongoloids: Their Contributions to theHistory & Culture of India.* Calcutta: The Royal Asiatic Society

Devkota, Kamal Raj. 2018. Navigating exclusionary-inclusion: Experience of dalit school children in rural Nepal. Globe: *Journal of Language, Culture and Communication*. Vol. 6, pp. 106-120. Alborg: Alborg University Press.

Grierson, George A. 1909. *Linguistic Survey of India* (vol. 3). Calcutta: Office of the Superitendent of Government Printing, India.

Kansakar, Tej R. 1999 The Newar Language:A Profile.*A Journal of Newar studies* 1:1,11-28,Portland USA

Kansakar, Tej R. 2011 A Sociolinguistic Survey of Newar (Nepal Bhasa) A report submitted to LINSUN,Central Department of Linguistics,TU

Kroskrity,P.V.2004 Language Ideologies.In A. Duranti(Ed.) *A companion to linguistic anthrolopology* (496-517) Oxford.Blackwell.

Malla,Kamal P. 2015.*From Literature to Culture:Selected writings on Nepalese studies,*1980-2010 Kathmandu Social Science Baha.

Nepali, Gopal Singh. 1965. *The Newars: An Ethno-Sociological Study of a HimalayanCommunity.* Bombay:United Asia Publications.

Regmi, D.R. 1960 /1969. *Ancient Nepal.*Calcutta: Mukhopadhyaya. Third Edition.

DETERMINING OFFICIAL LANGUAGES IN THE FEDERAL STATES IN NEPAL

Dan Raj Regmi

While population is the major criterion for determining official languages, other criteria such as language vitality, identity, linguistic right, linguistic geography should be rationally evaluated along with population in terms of numeric rating scales. Exclusively based on merit list, official language should be recommended in the federal states in Nepal.

1 Background

This paper presents some elementary techniques for assessing the criteria for determining official languages other than Nepali in the federal states in Nepal and suggests some strategies for meeting the challenges to be faced in the implementation of those techniques in Nepal.

Prior to consolidation in the eighteenth century, local forms of speech were uninterruptedly recognized as official languages in the twenty-two and twenty-four petty states of Nepal. The foundation of linguistic pluralism deep-rooted in Nepal, in a planned way, began to be traumatized with the recognition of Gorkha Bhasa (Nepali) as the only official language of the consolidated Nepal. This plan of traumatization effectively prolonged in the Rana Period (1846-1950) was forcefully accelerated in the Panchayat Period (1960-1990). Language assimilation policy reverberated in "One nation, one language", founded in monolingual ideology, was uninterruptedly enjoyed by the state until the promulgation of the Constitution of Nepal in 1990. Nepali written in Devanagari script was declared as the language of the nation (i.e., official language) and all the forms of speech spoken as mother tongues as the national languages. Local Self Governance Act (2055 BS) authorized the local bodies to use local languages as official languages (Clauses 96 and 189).The endeavor of using Maithili and Newar in the local levels was crushed by the verdict of Supreme Court in 1999. However, the struggle continued in

its pace and strength.

The interim constitution of 2007 declared all the languages spoken as mother tongues as languages of nation and granted right of the use of mother tongues as official languages in the local bodies and government offices. However, no crucial attempts were made for the implementation of such provision in practice. The constitution of Nepal, 2015 framed under the fundamental principles of socialism, federalism and inclusion has, knowingly or unknowingly, made some "tricky" of biased provision for the use of national languages in the federal states other than Nepali. Some discourses for looking for criteria for determining the official languages in the federal states has been duly initiated by Language Commission formed in 2016. However, such discourse for proposing criteria needs to be further widened and academically rectified.

This paper is organized into nine section sections. In Section 2, we briefly present the theoretical underpinning of the paper. Section 3 explains the need for determining the criteria for selecting official languages. In section 4, we briefly discuss the efforts for determining criteria. Section 5 looks at the criteria for determining official languages whereas in Section 6, we propose the model for evaluation of the criteria. In Section 7, we discuss the major problems in implementing the criteria. Section 8 suggests some major strategies for solving the problems implementing the criteria. Section 9 concludes the paper.

2 Theoretical underpinning

Fairclough (1989:1) affirms that, in modern society, use of language may maintain and change power relations. Moreover, the power and ideology of the speakers of certain language determines the social and legal status of the language in concern. Selection of particular language/s as official language/s (i.e., language/s having specific legal status and used within

government (viz., in courts, parliament, and administration) in the whole country or specific areas/federal states) may contribute to unequal relation of power. As per the constitutional criterion of 'majority', only a few languages are likely to be recommended for official languages in the specific provinces. Minority languages (indigenous as well as symbol of ethnic identity) will never be designated as official languages. Such situation will further contribute to the domination of minority speech communities by 'majority' speech communities as Fairclough pointed out. In order to help increase consciousness among the minority speech communities about the possibility for their languages to be official languages in the federal states, indigenous languages have to be recommended, on merit basis, for official languages. The majority 'ideology' at any cost should not be promoted despite the fact that present constitution has not set any further provisions except 'majority criterion'.

3 Need for determining criteria

The Constitution of Nepal, 2015 has recognized all languages spoken as the mother tongues as the languages of the nation (Article 6) and the Nepali language in the Devanagari script as official language of Nepal (Article 7 (1)). Article 7(2) has granted states the right, by framing a state law, to determine one or more than one languages of the nation spoken by a majority of people within the state as its official language(s), in addition to the Nepali language. Article 7(3), has authorized the Government of Nepal to decide other matters relating to language on recommendation of the Language Commission. Undoubtedly, "majority" is the foundation of democracy. However, with a view to including the minority speech communities by considering the complex ethno-linguistic architecture of Nepal, apart from the major criteria of "majority" other realistic and sensible criteria have to be sought after and national consensus has to be maintained as soon as possible. Nepal presents a threefold ethnic/religious-linguistic structure (Proposal of LinSuN, 2008:14). Nepali, official language, is spoken as mother tongue by 44.6%, Maithili (11.7%), Bhojpuri (6.0%), Tharu (5.8%), Tamang (5.1%), Newar (3.2%), Bajjika (3.0%), Magar (3.0%), Doteli (3.0%), and Urdu (2.6%). Chhetri, the largest caste/ethnic group having 16.6% of the total population, is followed by Brahman-Hill (12.2%), Magar (7.1%), Tharu (6.6%), Tamang (5.8%), Newar (5.0%), Kami (4.8%), Musalman (4.4%), Yadav (4.0%) and Rai (2.3%) (Central Bureau of Statistics, 2012). This situation presents a dire need of determining the other criteria in consonance with the ethno-linguistic situation of Nepal. Further, they have to be rationally synthesized.

4 Efforts for determining criteria

The Language Commission conducted workshops and seminars participated by concerned stakeholders, speech communities and experts in different provinces and centers to determine the criteria as per its five year roadmap. The Commission has collected a number of suggestions for setting up the criteria for determining official languages in the federal states in Nepal. Some suggestions have also been collected while conducting sociolinguistic survey of some languages and dialects spoken in Nepal with the financial support of the Language Commission.

It is not an easy task to propose undisputed criteria for determining for determining official languages in the federal states in Nepal. However, six basic criteria, for the first time, were proposed in (Bandhu et al., 2017). They include population, multilingualism and identity, accessibility and language right, writing system and languages only in spoken form, availability of literature and use of the language in communication. Moreover, Bandhu et al. (2017) has strongly noted that the "Majority" criterion is not justifiable to the minority speech communities of Nepal. It has further strongly suggested that the basic criteria should be systematically synthesized. Bandhu et al. (2017) was immediately followed by Yonjan-Tamang (2017). The latter has proposed five criteria for determining the official languages in the federal states of Nepal. They include population, linguistic geography (continuity of habitation), local identity and historicity of language, socio-cultural identity, language vitality

and power. It is to be noted here that there are fundamental similarities between these two proposals. However, the former has laid primary emphasis on the spirit of multilingualism and identity whereas the latter has prioritized the number of speakers. Whatever criteria are proposed, they are virtually rooted on the national context and international practices exercised mainly in India, Pakistan, China, Russia, Papua New Guinea, Latvia, USA(Hawaii).

In India, Hindi (written in Devanagari script) spoken by 14.5 to 24.5% as mother tongue, people have accepted it as the official language considering the principles of majority, identity and national integrity. Besides Hindi, English is also acknowledged as the official language mainly on the principle of national integrity. In Pakistan, 7.59 % speak Urdu as mother tongue and Urdu has got the status of official language on the principles of national identity and language of wider communication rather than majority criterion. China has accepted 'Standard Chinese', one of the dialects of Mandarin Chinese, as the official language of the country.

Russia, a multilingual country, has recognized only Russian as the official language at central level. However, in the provincial level, 35 mother tongues including Russian have been approved as official languages. In Papua New Guinea, more than 850 languages are spoken. However, only four languages, viz. Tok Pisin, Hiri Motu, English and the Sign language are recognized as official languages. In Latvia, Latvian is the official language constitutionally. A noteworthy matter in Latvia is that Russian spoken by around one-fourth of the population has been deliberately categorized as a foreign language. In the USA, languages spoken by minorities on the principle of inclusion have been recognized as the official languages in the state level. In Hawaii, English and Hawaiian are the official languages, whereas in Alaska, around twenty mother tongues including English have been approved as official languages.

Regmi (2017b) has extended the criteria for determining the official languages proposed in Bandhu et al. (2017) and briefly evaluated each basic criterion in terms of numeric rating scales. It has proposed for making a merit list of all the

languages before recommending for the official languages in federal states. In Nepal, indeed, majority, identity, national integrity, language of extensive communication, policy of inclusion may, broadly, be the major criteria for determining the official languages at the provincial levels.

5 Criteria for determining official languages

It is not an easy task to maintain unanimous consensus in setting criteria for determining official languages in the federal states in Nepal. However, consensus has to be sought after following discussions and interactions among stakeholders, speech communities and language experts. The criteria so far to be proposed are naturally liable to be revised as far as further ethno-linguistic information is made available. Some criteria may be proposed on the basis of international practices, interactions and seminars. They are briefly discussed as follows:

a. Number of speakers

A language or languages spoken by a specified percentage at the provincial or local level may be taken as one of the criteria for determining the official status of the language(s) in Nepal. However, such criterion may not justifiable as there has not yet been conducted a reliable census of the languages in Nepal (Regmi, 2018). Thus, other criteria such as multilingualism and identity have to be considered along with the criterion of number of speakers.

b. Language vitality

Normally, a language or languages which is/are spoken in all most domains of language use by all age groups and learnt as the first language/s should be chosen as official language/s at the local or provincial levels. Endangered, shifting, severely endangered languages should not be prioritized at any pretence.

c. Ethno-linguistic identity

In the present context of Nepal, language is a symbol of ethnic identity. Thus, while choosing a language or languages for official purpose, historicity and social-cultural identity have to be considered as basic criteria.

d. Accessibility and linguistic right

In consonance with the principle of multilingualism, the minority languages have to be prescribed for official use for providing the equal access to the government services and facilities.

e. Language of wider communication

A language or languages functioning as lingua-franca in the provincial levels has/have to be recommended for official language/s.

f. Linguistic geography

A language spoken in densely populated communities has to be recommended as official language at the provincial levels.

g. Linguistic originality (origin of language)

For promoting the national integrity, languages originated or evolved in the native land have to be recommended. Such languages may be useful tools for transmitting the life-crucial knowledge and strengthening the national identity.

h. Development of writing system

The official language is mostly used in documenting official records in written form. Thus, languages without written tradition should not be preferably recommended for such purpose.

i. Linguistic material development

Normally standardized form of a language is used in government offices. Thus, the languages that have standard grammars, dictionaries and textbooks are the right candidates for the recommendation for official language.

j. Corpus development

Standard grammars, dictionaries and textbooks can be easily prepared for the languages with available corpus. Therefore, the languages that have already developed corpora should be prioritized for official use.

k. Availability of literature

Literature makes a language rich. A language with a rich literature may be easily and effectively used in different levels of education in the country. Thus, availability of literature should also be considered as the basic criterion for recommending a language for official language in the federal states in Nepal.

6 Evaluation of the criteria

In Section 4, eleven criteria have been proposed for determining official languages in the federal states in Nepal. These criteria have to be systematically synthesized for the proper implementation. However, such a task is not easy for mainly two reasons. First, all criteria are not equally important in the context of Nepal. Second, undisputable model for evaluating criteria is not available in the literature. One of the models of practical rating scale is numeric rating scale. It is a very difficult task to synthesize all the criteria so far proposed in Section 4.

Table 1: Evaluating the basic criteria in terms of numeric rating scales

Major criteria	Numeric rating scale				
	1	2	3	4	5
Population				x	
Language vitality			x		
Ethno-linguistic identity					x
Accessibility and linguistic right					x
Linguistic geography		x			
Linguistic originality			x		
Development of writing system				x	
Linguistic material development				x	
Corpus development			x		
Availability of literature		x			

[1= poor, 2= fair, 3= good, 4= very good and 5= excellent]

It is to be noted that all criteria are not equally important in the context of Nepal and they have to be synthesized by using a practical rating scale. As per this rating scale, each criteria may be

evaluated in terms of five rating scales (viz., 1= poor, 2= fair, 3= good, 4= very good and 5= excellent). Table 1 presents a model of evaluating the basic criteria in terms of numeric rating scales.

Table 1 indicates that a language (X) proved to be excellent in all basic criteria in terms of numeric rating scales may be awarded 50 marks at the maximum. Thus, 50 may be supposed to be the full marks in the evaluation. However, in Nepal, a language (X) may be awarded 35 at the maximum. In Nepal, linguistic right of the minority speech communities has to be ensured in the spirit of multilingualism. Besides, migration rate has to be maintained as minimum as possible to encourage unified settlement in Nepal. Recognizing ethnolinguistic identity is also very important factor. As we have a number of cross-border languages such as Maithili, Bajjika, Bhojpuri, Awadhi, Rajbansi and Tibetan, linguistic originality has also been considered in Nepal. Development of writing system is awarded only three as there are a number of pre-literate languages in Nepal.

Table 2: Basis of evaluation of criteria

Major criteria	Rating scale	Basis for evaluation
Population	4	To ensure linguistic right to minority speech communities
Language vitality	3	To encourage endangered and shifting languages
Ethno-linguistic identity	5	To recognize ethnolinguistic identity
Accessibility and linguistic right	5	To ensure linguistic right in the spirit of multilingualism
Linguistic geography	2	To encourage unified settlement
Linguistic originality	3	Existence of cross-border languages
Development of writing system	3	Prevalence of pre-literate languages
Linguistic material development	4	Prevalence of a little linguistic materials
Corpus development	2	Prevalence of a little corpus development
Availability of literature	3	Prevalence of unavailability of literature

Table 2 presents rating scale of each basic criterion along with the rationality behind the particular rating.

It is to be noted that the bases of the evaluation of basic criteria are rooted on the present ethno-linguistic complexity of Nepal, situation of language vitality, statistics of speakers of languages and their distribution, issues of ethno-linguistic identity and principle of inclusion.

Based on the numeric rating scales given in Table 1, each language spoken in the areas where provincial and local levels are located has to be evaluated in overall. Each basic criterion, in turn, has to be turned into five rating scales. A language (X) may be awarded five marks at the maximum provided it is spoken by 5% or above. Likewise, a language spoken by more than 4% and spoken by less than 5% may be awarded 4 marks at the maximum. Similarly, a language spoken by more than 3% and spoken by less than 4% may get only 3 marks.

In this model, some marks have to be awarded to a language spoken by less than 1%. Table 3 presents a model for evaluating the number of speakers on the basis of numeric rating scale.

Table 3: Model for evaluating the number of speakers on the basis of numeric rating scale

Basis for evaluation	Numeric rating scale				
	1	2	3	4	5
The language spoken by 5% or above					x
Language spoken by more than 4% and spoken by less than 5%				x	
Language spoken by more than 3% and spoken by less than 4%			x		
Language spoken by more than 2% and spoken by less than 3%		x			
Language spoken by more than 1% and spoken by less than 2%	x				

Likewise, while evaluating the criteria of language vitality, a vigorous language may be awarded five marks and a critically endangered 1 mark only.[1] Table 4 presents a model for evaluating language vitality on the basis of numeric rating scale.

Table 4: Model for evaluating language vitality on the basis of numeric rating scale

Basis for evaluation	Numeric rating scale				
	1	2	3	4	5
Vigorous					x
Endangered				x	
Shifting			x		
Severely endangered		x			
Critically endangered	x				

Similarly, in the evaluation of ethno-linguistic identity, one ethnic group-one language will be awarded five marks whereas the speech community in which language of wider communication is used more than the mother tongue will receive only one mark. Table 5 presents a model for evaluating ethno-linguistic identity on the basis of numeric rating scale.

Table 5: Model for evaluating ethno-linguistic identity on the basis of numeric rating scale

Basis for evaluation	Numeric rating scale				
	1	2	3	4	5
one ethnic group one language					x
one ethnic group but more than one dialect				x	
one ethnic group but more than one language			x		
ethnic group using mother tongue and LWC equally		x			
ethnic group using LWC more than mother tongue	x				

In this way, all the basic criteria have to be systematically evaluated. We have to prepare a merit list of all the languages spoken in the areas located local and provincial levels. We have to follow the following procedure for preparing merit list.

A language (X) spoken by more than 5% or above may obtain five marks in maximum. In the context of Nepal, as given in Table 1, language X can obtain only four marks. Thus, language X is awarded only four marks in the final evaluation. Similarly, language X obtained four marks in the

evaluation of language vitality as being endangered. However, as given in Table 1, the maximum marks to language vitality have been conferred only three marks. Thus, at the end, language X can obtain only 2.4 marks.

The formula may be written as follows:

$$\frac{\text{Marks obtained (4)}}{\text{Max. Marks (5)}} \times \text{Max. marks for LV (3)} = 2.4$$

In this way, we have to prepare a merit list on the basis of the grand total obtained by X, Y, Z languages spoken in the federal states of Nepal. The languages falling in merit lists should not be recommended for official languages. The languages that do not qualify for the eligibility of official status need to be strengthened so that they can compete next time. This process should be continued.

7 Major problems in implementing the criteria

Nepal requires to assuring the linguistic rights of the speech communities in consonance with the spirit of multilingualism. Thus, the criteria for choosing the official languages have to be decided on the basis of reliable and full information about the languages of Nepal. Such decision should be rooted on the principles of inclusion and ethno-linguistic reality of Nepal in order to enable the people to rightly use their linguistic rights. Speech communities in Nepal have not yet been made conscious about the interrelationship between the language, culture and society. The numbers of languages, their speakers as well as the distribution enumerated in different censuses of Nepal have not been fully reliable. The study of languages in Nepal is in its infancy. The studies done so far have to be corroborated with present situation of languages in Nepal. The major problems are briefly discussed as follows:

a. Lack of the language census

Till today, there has not been conducted scientific census of the languages of Nepal. Thus, Nepal lacks reliable and undisputed statistics as to the number of languages, their distribution and speakers. Languages counted by enumerators with limited knowledge of language situation of Nepal

and limited number of questions in 2011 Census have number of problems (Regmi, 2018).

There is no uniformity in the number of languages enumerated in the six censuses carried out between 1992/54 to 2011. In 1952/54 census, 52 languages were enumerated as the mother tongues of Nepal. In the following three censuses, viz. 1961 (36), 1971(17), 1981(18) and 1991 (31) the number of languages is haphazardly and unreliably enumerated disgracing the dignity of speech communities of Nepal. After the reinstatement of democracy, in two censuses, viz., 2001(92) and 2011(123), the number of languages has increased. Heritage languages of ancestors or language of ethnic identity has been enumerated as mother tongues even though informants do not speak except Nepali.

Many international languages including English, French, and Spanish have been enumerated as mother tongues of Nepal. Some languages have been enumerated as two different languages. There are a number of inconsistencies as to the distribution of the speakers of the languages. Around 24 languages belonging to Rai-Kirati group have been enumerated as independent languages. At the same time, Rai with more than fifteen thousand speakers is again separately enumerated.

In reality, the speakers of Nepali as the first language are gradually increasing as many speech communities are shifting to Nepali. However, 2011 Census unreliably records 44.6% of population being the speakers of Nepali as mother tongue in Nepal. Thus, it would be quite impossible to ensure linguistic rights of the speech communities considering the number of speakers as the major criterion for determining official languages in Nepal.

b. Lack of a detailed study of vitality of language

Based on the Expanded Graded Intergenerational Disruption Scale model proposed by Lewis and Simons (2010) Regmi (2013; 2017a) has made a preliminary assessment of the vitality of the languages enumerated in 2011 Census.[ii] This assessment reveals that only less than 44% (53) languages are vigorous /safe. More than 41% (51) languages are threatened, i.e., only child

bearing generations are transmitting these languages to their children. Likewise, 8.9% (11) languages are shifting and 4.87% (6) are moribund. Similarly, 0.8% (1) is nearly extinct and 0.8% (1) dormant. Generally, more than 56% of the languages are facing different levels of language endangerment in Nepal. As per the criteria proposed by Lewis and Simons (2010), a language used orally by all generations and being learnt by children as their first language is considered as the vigorous language. Vulnerable/threatened (i.e., languages used orally by all generations but only some of the child bearing generations are transmitting it to their children) and definitely endangered/shifting (i.e., the child-bearing generation knows the language well enough to use it among themselves but none are transmitting it to their children) have to be promoted as vigorous languages if they are to be recommended as official languages in federal states in Nepal. Similarly, unless severely endangered/moribund (i.e., the only remaining active speakers of the language are members of the grandparent generation) and critically endangered/nearly extinct (i.e., the only remaining speakers of the language are members of the grandparent generation or older who have little opportunity to use the language) are revitalized and upgraded to the vigorous level, they may not be practically prescribed for the official languages. A detailed vitality study is also required for eliminating the inconsistencies occurred in the census of the languages.

c. Lack of dissemination of the sociolinguistic situation of the languages

Linguistic Survey of Nepal (2009-2017) has already revealed the sociolinguistic situations of around 96 languages/dialects of Nepal. However, they have not yet been properly disseminated among the speech communities. Without the proper knowledge of the domains of language use, multilingualism, dialectal variations, language attitude, language resources and aspirations of the speech communities for the development of their language, no effective planning may be prepared to work with the spirit of multilingualism in Nepal. Such planning should give emphasis to the use of indigenous languages in different domains of language use.

d. Lack of the study of accessibility of the speech communities

2011 Census provides some preliminary information as to the accessibility of different castes/ ethnicities on the principle of inclusion. However, there has not been yet conducted the survey of accessibility of the facilities of linguistic communities in Nepal. Besides, there is a lack of proper information on ethnicity/caste and language in Nepal. According to 2011 Census, around 10 castes (Kshetry, Bramins (Hills), Kami, Thakuri, Sarki, Sanyasi, Gaine, Badi, Khawas, Damai/Dholi) speak Nepali as the common mother tongue whereas 53 ethnicities have their different mother tongues. Thus, the survey of the accessibility of different castes/ ethnicities cannot provide a clear picture of the accessibility of the peoples and linguistic rights.

e. Lack of writing system

Only a few languages of Nepal have a tradition of writing. Such languages include Nepali, Maithili, Tibeti/Sherpa, Tamang, Newar, Limbu, Bhojpuri, Awadhi and Lepcha. Many languages of Nepal are preliterate. Normally, there is a tradition of selecting languages with established writing traditions. Generally languages in which grammars, dictionaries as well as textbooks have been made available are recommended for official status. Thus, it would be practically very difficult to select preliterate languages for such purposes. However, a preliterate local language may have to be recommended for official use.

f. Lack of linguistic research

Only a few languages of Nepal have been studied comprehensively. Such languages may include Nepali, Newar, Tamang, Maithili, Bhojpuri, Awadhi, Limbu, Bantawa, Koyee, Dumi, Bhujel, Sunar, Thami, Magar Kaike, Dolkha Newar, Kham, Darai, Hayu, Belhare, Athpahare, Chamling, Yakkha, Kusunda, Raji, Bajhangi, Achhami, Dadeldhuri, Rajbansi and Chepang. Many languages are still looking forward to be comprehensively studied.

g. Lack of consciousness about the importance of mother tongue

Michael Krauss (2002), a linguist, rightly remarks that any language is a supreme achievement of a uniquely human collective genius. He further points out that any language is divine and endless a mystery as a living organism. By far the most important remarks he has made is that language death matters, i.e., "the extinction of any animal species diminishes our world, so does the extinction of any language." Nelson Mandela, a political leader, once upholding the significance of mother tongue remarked that the address made in mother tongue only can touch the heart of addressee. He notes "If you talk to a man in a language (i.e., other than mother tongue), that goes to his head. If you talk to him in his language (i.e., mother tongue), that goes to his heart." For the great dramatist of Nepal, Balakrishna Sama, any language is civilization, and the symbol of every progress, i.e., language makes immortal every victory and every success. Due to the lack of education and knowledge about the importance of language, the attitude towards the mother tongue among the speech communities has not yet been effectively promoted as expected.

8 Major strategies

Determining criteria for the official languages in Nepal has been a very complex task. Unless short term and long term language planning with specific strategies is framed for the preservation, promotion and development of the languages of Nepal, no criteria would bear the expected result. There are some specific strategies to solve the problems in the implementation of the criteria. They are as follows:

a. If criterion of majority of speakers stated in the constitution is admitted, a systematic and scientific census of languages should be immediately conducted. Uncertainty about the number of languages, their speakers as well as the places of the speech community can be eliminated.

b. The vitality of the languages of Nepal has to be assessed following the model for vitality assessment referred to as Expanded Graded Intergenerational Disruption Scale expanded and revised by Lewis and Simons in 2010.

c. Ethno-linguistic survey has to be immediately conducted for the ethno-linguistic identity of the speech communities in Nepal.

d. Language documentation programs have to be immediately framed and executed for strengthening the seriously endangered languages and assuring the accessibility of such speech communities in the services and facilities provided by the governments.

e. There should be the survey of accessibility of the facilities of linguistic communities in Nepal.

f. The major findings of the sociolinguistic survey of the languages of Nepal as to the domains of language use of all languages and dialects, situation of multilingualism, dialectal variations, language attitudes, language resources and aspiration of the speech communities for language development have to be duly considered while making any crucial decision about languages of Nepal.

g. An action plan for the writing of unwritten languages has to be immediately prepared and executed with the active participation of the speech communities and linguists.

h. Research of languages with active participation of speech communities has to be carried out.

i. Awareness program in the speech communities has to be framed and executed effectively and immediately.

9 Summary and conclusion

The Language Commission is assigned a special task to determine the criteria of selecting one or more languages as official languages in the federal states in Nepal. Some preliminary discussions have been made among stakeholders, speech communities, linguists and concerned authorities as to such criteria. However, right now, it is not easy to set "unanimously agreed criteria" in consideration to the language composition, federal structure and ethnolinguistic complexity in Nepal. Internationally, some criteria such as population, multilingualism, language of wider communication, national identity and national integrity have been in practice for assigning the role of official language. Till the date, identity and multilingualism and population are the competing criteria. However, in the present context of Nepal, all the major criteria have to be evaluated by applying the numeric rating scale. The marks obtained in individual criteria have to be re-evaluated. On the basis of the grand total of the marks, a merit list has to be prepared of the languages of each state. Language Commission is advised to recommend the government to assign the role of official language on the basis of the merit. No doubt, setting criteria has been a complicated task thanks to the lack of reliable number of speakers, lack of writing system and lack of the study of the languages of Nepal. Such hitches can be mitigated by framing an effective language planning immediately. Besides, census of languages and assessment of the vitality of all the languages of Nepal have to be immediately carried on. Recommending deserving languages by setting reliable and effective criteria to be the official languages in the federal states is the dire need of today's Nepal.

References

Bandhu, Churamani; Dan Raj Regmi and Prem Phyak. 2017. Sarakari kamkajko bhasa nirdharanaka adharharu (Criteria for determining official languages). A paper presented at the symposium on Constitutional Responsibility of Language Commission and its Future Roadmap organized by Language Commission, July 12, 2017, Kathmandu.

Central Bureau of Statistics. 2001. *Population Census*, Kathmandu: National Planning Commission.

Central Bureau of Statistics. 2012. *National Population and Housing Census 2011: National Report*. Kathmandu: National Planning Commission.

Fairclough, Norman. 1989. *Language and Power*. New York: Longman.

Government of Nepal. 2015. *The constitution of Nepal*. Kathmandu: Ministry of Law, Justice, and Parliamentary Affairs.

Lewis, M. P. and G. F. Simons. 2010. Assessing Endangerment: Expanding Fishman's GIDS. *Revue Romaine de Linguistique* 55.103-120.

Linguistic Survey of Nepal. 2008. *Proposal of Linguistic Survey of Nepal*. National Planning Commission, Nepal.

Regmi, Dan Raj. 2013. Nepalka Bhasaharuko Sarvekshan: Sthiti ra Chunauti (Linguistic

survey of the languages of Nepal: Situation and challenges). A paper presented at the national workshop on the development of language policy in Nepal organized jointly by Ministry of Education and Nepal Academy, July 10, 2013, Kathmandu.

Regmi, Dan Raj. 2017a. Convalescing the endangered languages in Nepal: Policy, practice and challenges, *Gipan*, Vol.3.1, January, 2017.

Regmi, Dan Raj. 2017b. bhasa chanotka jatilata (Complexity in selecting languages). *Kantipur Daily*. 2 September, 2017 (Bhadra 17, 2074).

Regmi, Dan Raj. 2018. tapainle bolne dosro bhasa kun? (What language do you speak as the second language?). *Kantipur Daily*, 29 August, 2018 (Bhadra 13, 2075).

Yadava, Yogendra P. 2013. Linguistic context and language endangerment in Nepal. *Nepalese Linguistics,* Vol.28, 260-272.

Yadava, Yogendra. 2015. Language use in Nepal. *Population Monograph of Nepal. Vol. II (Social Demography)*. p. 51-64.

Yonjan-Tamang, Amrit. 2017. Sarakari kamkajko bhasa nirdharana (Determining official languages). A paper presented at the interaction program on Language Policy in the Constitution of Nepal and Way forward organized by National Foundations for Development of Indigenous Nationalities, August 2, 2017, Lalitpur.

---

[i] See section 6(b) for the definitions of different levels of language vitality.

[ii] See Yadava (2013) and (2015) for Linguistic context and language endangerment in Nepal language use in Nepal, respectively.

A PRELIMINARY STUDY OF MNAR

Ruth Rymbai and Arvind Kumar Rawat

The present paper is an initiatory investigation of Mnar, a dialect of Khasi, which is classified under the Mon-Khmer branch of the Austro-Asiatic language family. Mnar occupies the area known as Jirang, which falls under the RiBhoi district of Meghalaya. Although, Mnar shares common structural traits with other Austro-Asiatic languages, it still exhibits unique properties, providing a great opportunity for succinct investigation of the structural patternsexisting in the language.

Keywords: Mnar, Mon-Khmer, word order

1 Introduction

Diffloth (2005) coined the term 'Khasian' to refer to the varieties of Khasi spoken in the Khasi and Jaintia hills of Meghalaya. As per the Census of India (2011), the total population of Jirang is 30,919. Mnar is not well studied linguistically and there is hardly any account of studies pertaining to Mnar in the field of literature. Further, Gurdon (1914) has listed Mnar as one of the Khasi dialects and mentioned briefly about the numeral system in Mnar. Although, he listed Mnar as one of the Khasi dialects, he still mentions that Mnar is affiliated to Synteng, Lakadong and Amwi rather than to Standard Khasi. As a result, there is no mutual intelligibility between Khasi and Mnar.

The aim of this paper is to provide a brief outline with a view of arriving at an understanding of the nature of structural constructions in Mnar and to perceive how such analysis can enrich the knowledge of fairly unknown languages. The components chosen are intended to enable a comprehensive view of the structural attributes of Mnar. The fact that there is very little work available that compares and contrasts the rich assortment of data found in Mnar, the empirical evidence is drawn from primary sources, i.e. field work and language consultants. It should also be added here that attention is centered on the formal aspects of typology.

2 Data

The data collection for this paper was undertaken in the village of Jirang. The data analysis comes from fieldwork transcription in the form of word list, derived from Swadesh List. For sentences the consultants were asked to translate Khasi texts into Mnar, so that the investigators could note down and transcribe the spoken data. The drawback of the present study is that it only presents data elicited by way of translation. Prospective work will need to include naturally occurring data related through folktales, narratives and communicative events that people regard as important parts of their cultural heritage, occurring in oratory, song, dance recitals of poetry or story-telling, rituals, and so on. Unfortunately, no audio or video recordings were made in this fieldtrip.

3 Phonology

The segmental phonology of Khasi and few other dialects including Pnar, Lyngngam and Bhoi has been presented adequately by Nagaraja (1985, 1993 1996, and 2014) and Ring (2012, 2014, and 2015). However, as mentioned by Diffloth and Zide (1992), the "so called Khasi dialects, such as Synteng (Pnar), Lyngngam and Amwi (War) are clearly distinct but related languages" (cited in Koshy and Wahlang 2011). Daladier (2011) in an extensive research on Pnar, War, Lyngngam and Standard Khasi also recognizes the fact that the aforesaid varieties are mutually unintelligible and, therefore, adduced that these varieties are four disparate languages. Accordingly, these varieties exhibit certain distinctive geographical and structural features. In keeping with the typical typological phonological lineation of Austro-Asiatic (AA) languages, Mnar exhibits the following characteristics.

3.1 Consonants

Mnarhas five places of articulation of stops bilabial /p,b)/; alveolar /t,d/; palatal /c,ɟ/; velar/k/ and glottal /ʔ/ and four for nasals: bilabial /m/; alveolar /n/, palatal /ɲ/ and velar /ŋ/. The fricatives available are /s/ and /h/. Furthermore, Mnar has one rhotic /r/, one lateral approximant /l/, and two semi vowels /w and j/.

Additionally, Mnar has two stop series, usually distinguishing between voiceless (unaspirated) and voiced stops /pʰ, bʰ, tʰ, cʰ and kʰ/. Aspiration is phonemic, thereby giving way to three way voicing contrast in oral stops. Nagaraja (1996) found the occurrence of voiced velar stop /g/ in his study of Lyngngam, but in Mnar its occurrence is limited and at the same time suspicious. First and foremost it occurs as an alternate variant of ka '3SF' [ka~ga] and ki '3PL' [ki~gi]. Secondly, there is no phonemic contrast for /g/ and that its occurrence is limited to the initial position only as in gana 'this'; gatai 'that'; gitai 'those'; gira 'relative marker' etc.

The following minimal and sub minimal sets illustrate the contrast between Mnar consonants:

Table 1: Consonant phoneme opposition in Mnar

Stops	
/p/ pɔʔ 'belly'	/b/ bɔʔ 'to tie'
/t/ tɛw 'tree'	/d/ dɛw 'cloth'
/c/ cut 'choose'	/ɟ/ ɟut 'count'
/k/ ksei 'dance'	/p/ psei 'sister-in-law'
Aspirated vs. Unaspirated	
/p/ pi '2PL'	/pʰ/ pʰi 'ten'
/b/ biʔ 'poison'	/bʰ/ bʰiʔ 'good'
/t/ tai 'that'	/tʰ/ tʰai 'herd'
/c/ ci 'see'	/cʰ/ cʰim 'blood'
/k/ kam 'work'	/kʰ/ kʰam 'handful'
Trill vs. Lateral	
/r/ reʔ 'auxiliary verb'	/l/ leʔ 'past tense marker'
Nasals	
/m/	/n/

mar 'each'	nar 'iron'
/n/ nar 'iron'	/ɲ/ ɲaʔ 'drive'
/ɲ/ tʰaɲ 'red'	/ŋ/ tʰaŋ 'burn'
Fricatives	
/s/ saʔ 'stay'	/h/ ha 'locative case marker'
Approximants	
/w/ waʔ 'also'	/j/ jaʔ 'black'

As already mentioned, Mnar has four nasals; m, n, ɲ, ŋ. They can occur in both onset and coda positions as illustrated below:

Table 2: Distribution of nasals

Phoneme	Onset	Coda
m	mat 'eye'	crɲam 'green'
n	nɔ 'leg'	san 'five'
ɲ	ɲim 'cry'	tʰaɲ 'red'
ŋ	ŋa '1SG'	tɨrlaŋ 'ear'

The inventory of syllable-final consonants is smaller than that of onset consonants in Mnar in that there is only one series of stops, which are always voiceless and typically unreleased, i.e. accompanied by glottal restriction which stops the airflow (skep˺ 'ribcage'; sut˺ 'veins'). Other consonants that can occur in the coda position are ŋ (kṇrɔŋ 'long'); k (wak 'many'); r (i: r 'two'); ʔ (sniaʔ 'skin'); m (cʰim 'blood'); n (lmen 'tooth') and ɲ (tʰaɲ 'red').

3.2 Vowels

The richness of Mnar vocalic system is typically Austro Asiatic (Jenny et al. 2014), with a minimum of three degrees of vowel height. It has a total of 8 vocalic nuclei: one high front vowel (i), two mid front vowel series (e,ɛ), one low front vowel (a), three back vowels (u, o, ɔ) and one central vowel (ɨ).

Table 3: Vowel contrast pairs in Mnar

/i/ im 'sibling'	/a/ am 'progressive marker'
/a/ sam 'take'	/e/ sem 'bathe'
/a/ ma '2sm'	/u/ mu 'mother'
/ɛ/ ɛʔ 'loud'	/a/ aʔ 'auxiliary verb'
/e/ reʔ 'auxiliary verb'	/ɛ/ ɛʔ 'loud'
/e/ hen 'measuring unit'	/u/ hun 'child'
/ɛ/ ɛt 'stick'	/ɔ/ ɔt 'cut'
/i/ liʔ 'go'	/e/ leʔ 'past tense marker'
/a/ saʔ 'stay'	/ɔ/ sɔʔ 'fruit'
/u/ tʰuʔ 'search'	/ɔ/ tʰɔʔ 'write'
/i/ wiʔ 'earthworm'	/ɔ/ wɔʔ 'grandfather'
/o/ noʔ 'throw'	/ɔ/ nɔ 'leg'

As with theKhasian languages (Nagaraja 2014: Ring 2014 and 2015),vowel length contrast is irregular and restricted to a subgroup of vowel in Mnar. Only [a:] shows phonemic contrast in length. The long vowel [i:] is inconsistent as well. Other examples with long vowels include: pla:m 'cloud' sta:t 'wise' pa:m 'slice' i:r 'two' bi:m 'eat' sɟi: 'rice' ki:ʔ 'climb'.

Mnar is similar to Pnar (Ring, 2012 and 2015) in its inventory of diphthongs. There are only two phonemic diphthongs [ia] and [ɔu]. [ia] is found in closed syllables as in sniaʔ 'skin'; sʔiaŋ 'seed'; tʰia 'sleep'. While, [ɔu] is found in open syllables as in ksɔu 'dog'; smɔu 'stone'; hlɔu 'door'. Mnar also shows the presence of a vowel sound with either a labial [w] or palatal [j] glide, that either precedes (an on-glide) or follows (an off-glide) the main vowel.

Table 4: Mnar diphthongs with on- and off- glides

Diphthongs	Description	Word	Gloss
/ei/	a front mid vowel with palatal off-glide	tei	'hand'
/ai/	an open vowel with palatal off-glide	mai	'face'
/oi/	a back mid vowel with palatal off-glide	moi	'buffalo'
/ie/	a front mid vowel with palatal on-glide	biet	'fool'
/ui/	a front close vowel with labial on-glide	khuic	'clean'

3.3 Syllable structure

Phonological words in Mnar consist of monosyllabic, sesquisyllabic (Matisoff 1973:86) and multisyllabic roots. Disyllabic words are most frequent and phonologically they are merely the composite of two syllables, each of which follows the same phonological requisites. Trisyllabic words are rare, if found to prevail they are an outcome of word formation (compounding).A syllable in Mnar may consist of an initial consonant cluster (CC), an obligatory vowel nuclei V, and an optional final consonant phoneme (C). The formulaic structure of the syllable is (C_1) (C_2) V (C_3). The smallest word shape allowed in Mnar is a single vowel nuclei and the largest word shape consists of a complex onset of two consonants, a diphthong nucleus and a coda consonant.

Table 5: Structure of syllables in Mnar

Monosyllables	Word	Gloss
V	i	'1PL'
VC	im	'niece'
CV	mu	'mother'
CVC	san	'five'
CCV	sŋi	'house'
CVV	lei	'three'

3.3.1 Consonant cluster

The combination of consonants in the initial position is rich in Mnar, usually all consonants can occur, some of them go against the sonority ranking, but sequences of the same place of articulation are avoided. Mnar onset clusters

include the sequences of stop plus nasal, stop plus liquid, fricative plus stop, fricative plus nasal, liquid plus nasal, etc.

Table 6: Permissible onset clusters

Cluster	Word	Gloss
pn	pnuʔ	'salt'
ps	psen	'snake'
pr	praʔ	'know'
pl	pla:m	'cloud'
bl	blaŋ	'goat'
tŋ	tŋam	'cold weather'
tl	tlɔt	'weak'
tʰr	tʰrei	'six'
tʰm	tʰmi	'war'
cr	crŋam	'green'
cʰl	cʰlam	'cold water'
ktʰ	ktʰar	'axe'
kl	klɔŋsnam	'heart'
ks	ksaŋbʰi	'good'
km	kmen	'happy'
kʰl	kʰlɔu	'head'
sn	sniaʔ	'bone'
sk	skɔu	'sit'
sm	smɔu	'stone'
sŋ	sŋeic	'stout'
st	sta:t	'clever'
hl	hlɔu	'door'

3.3.2 Sesquisyllabic structure

A distinct syllable type termed 'sesquisyllabic' (Matisoff 1973:86) is found to exist in many Austro-Asiatic languages (Jenny et al. 2014) wherein a disyllabic word consists of an initial unstressed syllable often called a minor (Henderson, 1952) or pre-syllable followed by a stressed full syllable (main syllable). In this minor or pre-syllable the nucleus of the syllable is occupied by either of these sonorant sounds (l,r,m,n, ŋ), in lieu of certain weak vowels and as such they carry the main weight of the first syllable in a disyllabic word. This phenomenon can be explained with the disyllabic word ɟyṇdaŋ

'neck'. The pre-syllable in this example is ɟyn, which is orthographically spelled as ɟyn,and transcribed phonetically as ɟṇ or ɟɨn. Here the vowel is eliminated or rendered weak and the following nasal /n/ takes possession of the nucleus position, thereby giving rise to a sesquisyllabic structure.

3.3.3 Suprasegmental

Since most Austro-Asiatic languages are sesquisyllabic, they tend to be strongly iambic, wherein a weak and unstressed syllable is followed by a strong and full stressed syllable (Jenny et al. 2014). Likewise, the stress in Mnar is always in the last syllable of the word. The pre-syllable always gets the reduced stress and transitional vowel.

4 Morphology

Mnar is isolating, in that it is extremely analytic with words consisting of a single morpheme constituting a separate word and independent grammatical words. However, no language is purely or predominantly of one type. Thus, Mnar is also characterized by some degree of agglutination, mainly, in its technique of employing affixes to be juxtaposed to root words. The addition causes no significant changes in the root and the different affixes are readily identifiable and easily segmented from the root and from one another as illustrated in the examples below.

(1) u   pitar   bi:m   sɔʔ   u
    3SM   PN   eat   fruit 3SM
    'Peter eats fruit.'

(2) ga sapʰi   pṇ-rɔʔ   kaha-rara
    3SFPN   CAUS-praise   3SFACC-REFL
    'Saphi praised herself.'

Sentence (1) illustrates how Mnar comes close to being an isolating type. Each word in the sentence consists of just a single and free morpheme. There are monosyllabic words both lexical and grammatical (function words) as in biːm 'eat', sɔʔ 'fruit' and u '3SM'. Each morpheme is invariable in that the words are strung together in a sentence but without change; thus, biːm 'eat' does not inflect to show person, number or tense.

In sentence (2) the word pn̩-rɔʔ 'cause to praise' consists of two morphemes pn̩- 'a causative suffix and *rɔʔ* meaning 'praise'. The boundary between these two morphemes in the word is clear-cut; moreover, the identification of morpheme in terms of their phonetic shape is also straightforward. The anaphoric expression ha-rara is a combination of two morphemes ha- 'accusative case' and rara 'personal reflexive', here too the boundary is clear-cut.

The lack of inflectional categories as attested in Mnar is compensated by the extensive derivational morphology, including class changing derivational prefixes, compounding and reduplication. Compounding is common, and can consist of two or more nouns, a sequence of a noun plus verb, or a combination of two or more verbs.

Table 7: Compound words

Combination	Compound	Gloss	Gloss 2
Noun+Noun	sʔiaŋ+tɨmpɔn	back+bone	'backbone'
Verb+Noun	don+burɔm	be+honour	'honourable'
Verb+Adjective	ksaŋ+kmen	feel+happy	'elated'
Noun+Verb	aːm+biaʔ	water+spit	'saliva'

Derivational morphology in Mnar is operational through affixation attaining nominalization and causativization processes. Normally verbs are nominalised by adding prefixes such as tʰei- (its function is to derive noun stems from verb and adjectives) in tʰie+sali (NOMLZ+lazy) 'sloth'. Additionally, Mnar also has an agentive nominalising prefix men- as well as a causativizing prefix containing the element pn̩- is attached to a simplex verb *jip* meaning 'die' as in pn̩-jip 'cause to kill'.

Reduplication is yet another common word formation process attested in almost all Austro-Asiatic languages. It appears in various manifestations either as full or partial reduplication (imra-imra 'there'; ju-ju 'nothing'; muɔn-muɔn 'slowly'; ɟɔkʰa-mɔkʰa 'some'). Adjectives, question words and adverbs all have reduplicated forms in the various Khasian varieties (Ring 2014 and Nagaraja 2014). This also holds true for Mnar, in that it uses reduplication to mark emphasis, change the grammatical category of words and intensify the meaning through complete repetition of adverbials, deictic words and question words. Full reduplication is constructed by repeating the element in an identical form whereas in partial reduplication a consonant is changed in the initial position in the repeated element.

5 Pronominal system and gender

On account of the absence of obligatory morphological markingof tense, agreement, number or any other morpho-syntactic category generally expressed by inflection, syntactic notions such as subject and object are purely based on syntactic criteria in Mnar. Having said so, Mnar conforms to Subject-Verb-Object basic word order and is characterized by a fairly rigid word order, wherein constituents cannot befreely moved from one position to the other. It has a rich set of functional words which mark the grammatical properties of phrases and clauses. It is a head initial displaying modified-modifier ordering and has a noun classifier system. Before describing the word order typology of Mnar it is important to discuss the pronominal and gender system.

Table 8: Pronominal chart of Mnar

Singular	Nominative	Accusative	Genitive/Possessive
First Person	ŋa 'I'	ha-o 'me'	ɟɔ-o 'my'
Second Person	ma(M) pa(F) 'you'	ma(M) pa(F) 'you'	ɟɔ-ma(M) ɟɔ-pa (F) 'your'
Third Person	u / ka~ga 'he' / 'she'	wei / kai 'him' /'her'	ɟɔ-wei/ɟɔ-kai 'his'/ 'her'

Plural			
First Person	wi 'we'	wi 'us'	ɟɔ-wi 'our'
Second Person	pi 'you'	pi 'you'	ɟɔ-pi 'your'
Third Person	ki~gi 'they'	kei~gei 'them'	ɟɔ-kei 'their'

Mnar has a rich pronominal system with two number distinctions: singular and plural. Feminine and masculine gender distinction is seen in second person singular ma (M) and pa (F). Further it must be pointed out that Mnar has a nominative-accusative form of alignment in terms of case marking and/or verb agreement, wherein the subject (S) of an intransitive verb has the same case as the subject or agent (A) of a transitive verb, contrasting with the patient (P) / object (O) of a transitive verb which gets coded differently. Third person pronoun /u/ and /ka/ can be labelled as portmanteau morphs for its multifunctional attribute. The said pronominals can function as gender clitics, personal pronouns as well as agreement markers in Mnar.

Gender distinction in Mnar is seen not only in pronouns but also in lexical nouns (nominals). In Mnar nouns cannot function without gender, it is obligatorily attached before a noun. Some animate nouns can be divided into female and male. The following are the gender clitics: ka~ga (feminine); u (masculine) and ŋa (neuter) as in:

(3) ga sara
     F   PN
     'Sara'

(4) u ɟɔn
     M   PN
     'John'

(5) ŋa hai
     N   thing
     'thing'

5.1 Verbal agreement

Agreement in Mnar is overtly realized between an NP (noun phrase) and a verb. The verb obligatorily agrees with 3rd person subjects NP in terms of person and number and gender but there is no agreement for non-3rd person subject. The subject is marked by the pronominal clitics placed after the verbal root, though the place of occurrence of agreement is not fixed. The pronominal agreement clitics have the same shape as personal pronouns, the third person masculine is marked by u (masculine) and ka (feminine) respectively in the singular, and by ki in the plural; when an animate noun stands as the subject NP, it agrees with the verb by its clitic form. Here, it must be mentioned that the agreement pattern in Mnar is post verbal, post adjectival and post sentential unlike Khasi which has a preverbal agreement. Furthermore, Mnar is also a pro-drop language, where agreement enables the subject to be dropped; pro is a covert nominative case pronoun occurring in the subject position of the finite clauses, showing a rich verb agreement.

(6) u   ɟɔn   ieid   u   ha
     3SM   PN   love   3SM   ACC
     ga meri
     F PN
     'John loves Mary.'

In sentence (6) the highlighted u marks agreement of the verb ieid 'love' with its subject u ɟɔn 'John'. The subject is third person singular masculine, which is also marked by the u directly preceding the head noun ɟɔn 'John'. Further, the agreement marker is located after the verb.

(7) sa ɟi   ki   (pro-drop)
     eat rice   3PL.
     'They eat rice.'

The subject can be dropped as in (7), because the subject agreement marker is coded on pronominal clitic ki '3PL' following the verb.

5.2. General word order

The unavailability of morphological markingof tense, agreement, number generally represented by inflection, syntactic notions such as subject and object are purely based on syntactic criteria in Mnar. Having said so,Mnar conforms to Subject-Verb-Object basic word order and is characterized by a fairly rigid word order, wherein constituents cannot befreely moved from one position to the other. It has a rich set of functional

words which mark the grammatical properties of phrases and clauses. It is head initial language displaying modified-modifier ordering and has a noun classifier system.

The word order characteristics discussed in this section will be based on the writings and research of Greenberg (1963) and Dryer (1992). At the surface level, Mnar seems to exhibit three different types of word order. They are: (a) SVO, (b) VSO, and (c) VOS.

The following is an illustration to show the basic word order of Mnar (SVO) in a simple declarative sentence.

(8) ga meri   tieŋ   haicbi:m   ka
    3SF PN   cook   food   3SF
    'Mary cooks food.'

Sentence (8) exhibits S-V-O word order where the subject/agent precedes the main verb; and the object/patient follows the verb. However, possible alternate variation orders also exist in Mnar as in the following examples. A further research on word order variations is needed as no conclusion has been arrived at as yet.

(9) ieid   ki   wei   (VSO)
    love   3PL   3SM(ACC)
    'They   love   him.'

(10) dɛʔ a:m wi i:rwat si siŋ (VOS)
     drink water 1PLDD twice QUANT *day*
     'We drink water twice a day.'

As predicted by the implicational universal (Greenberg, 1963) for verb medial language which includes Mnar the genitive follows the head noun.

(11) am   saʔ ka   i   sʰillɔŋ
     PROG live 3SF   LOC   PN
     'She still lives in Shillong.'

(12) ga im   ɟɔ   ga   meri
     3SF niece GEN   3SF   PN
     'Mary's niece.'

Subject-Verb

(13) u   ban   lhɔʔsari u
     3SM PN   laugh   3SM
     'Ban laughs.'

Subject-Verb-Direct Object

(14) u ɟɔn   dat   u   ha   ga
     meri
     3SG   PN   beat 3SG   ACC
     F PN
     'John beat Mary.'

Subject-Verb-Direct Object-Indirect Object

(15) u   ɟɔn le?lep ai ʃitʰi u   ha
     3SM PN finish give letter 3SM   ACC
     ga meri
     F PN
     'John gave the letter to Mary.'

Subject-Verb-Complement

(16) u ɟɔn   leʔ ɔŋ u ba ŋat
     3SM PN   PST say 3SM COMP AUX
     liʔ ŋa i   sŋi
     go 1SG LOC   house
     'John said that I should go to the house.'

Subject-Verb-Relative

(17) la:m   ka tarei   gara   aʔ
     bring   f   knife   rel   aux
     i   sŋi   tieŋ-ɟi
     loc room   cook-food
     'Bring the knife which is in the kitchen.'

5.4. Prenominal Modifiers

Demonstratives precede the head noun and classifiers and numerals precede the head noun.

(18) a. i:r   kʰlen aʔ   ksɔu   ŋa
         two   CLF AUX   dog   1SG
         'I have two dogs.'

     b. sɔu bei aʔ hun antʰei ki
         four CLF AUX child girl PL
         'They have four daughters.'

5.5 Postnominal modifiers

Adjectives follow the head noun

(19) ga sʔir   sŋaid
     F hen   fat
     'fat hen'

5.6 Order of auxiliary verb

The auxiliary verb precedes the main verb

(20) leʔ   tʰɔʔ   u   ga   ʃitʰi
     AUX   write 3SG   F   letter
     'Did he write the letter?'

6 Conclusions

This work is an attempt to make few pertinent and preliminary observations on the structural features of Mnar. The findings presented in this paper do validate the fact that Mnar does share a number of features with other Austro-Asiatic languages like Khasi and Pnar. However, as matter of course, Mnaralso has some peculiar characteristics of its own, thus demonstrating that each language retains its identity in spite of intense contact with other languages.

References

Comrie, Bernard. 1981. *Language Universals and Linguistic Typology*. Chicago: Chicago University Press.

Croft, William. 1990. *Typology and Universals*. Cambridge: Cambridge University Press.

Daladier, Anne. 2011. The group Pnaric-War-Lyngngam and Khasi as a branch of Pnaric. *Journal of the Southeast Asian Linguistics Society* 4(2). 169-206.

Diffloth, Gerard. 2005. The Contribution of Linguistic Palaentology and Austroasiatic. In Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, (eds.)*The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. London: Routledge.

Dryer, Matthew. 1992. The Greenbergian Word Order Correlations. *Language* 68. 81-138.

Greenberg, Joseph. 1963. Some Universal of Grammar with Particular Reference to the Order of Meaningful Elements. In Greenberg, Joseph., (ed).*Universals of Human Language*. Cambridge, Massachusetts: MIT Press 73-113.

Gurdon, P. R. T. 1914. *The Khasis*. London: Macmillan and Co.

Henderson, Eugenie. J.A. 1952. The Main Features of Cambodian Pronunciation. *Bulletin of the School of Oriental and African Studies* 14.159-174.

Mathias Jenny and Paul Sidwell, (eds.) 2014.*The Handbook of Austroasiatic languages.*Vol 2. Leiden: Brill

Mathias Jenny, Weber Tobias and Weymuth Rachel. 2014. The Austroasiatic Languages: A Typological Overview. In Mathias, Jenny and Paul, Sidwell. (eds.) *The Handbook of* Austroasiatic Languages 2. 13-143.Part I, Ch. 2. Leiden: Brill.

Matisoff, James. 1973.Tonogenesis in Southeast Asia: In Consonant Types and Tone. Larry Hyman, (ed.) *Southern California Occasional Papers in Linguistics 1*. Los Angeles: UCLA.

Nagaraja, K. S. 1985. *Khasi: A Descriptive Analysis*. PhD thesis, Pune: Deccan College.

Nagaraja, K. S. 1993. Khasi dialects, a typological consideration. *Mon-Khmer Studies Journal* 23.1–10.

Nagaraja, K. S. 1996. The status of Lyngngam. *Mon-Khmer Studies Journal* 26.37– 50.

Nagaraja, K. S. 2014. Standard Khasi. In Mathias Jenny and Paul Sidwell, (eds.) *The Handbook of Austroasiatic Languages* 2. 1143–1185. Part II, Ch. 19. Leiden: Brill.

Koshy, Anish and Maranatha G.T. Wahlang. 2011. MnarMorpho-Syntax: A Preliminary Study. *Indian Linguistics* 72: 153-171

Ring, Hiram. 2012. A phonetic description and phonemic analysis of Jowai-Pnar. *Mon-Khmer Studies,* 40:133–175.

Ring, Hiram. 2014. Pnar. In Mathais, Jenny and Paul, Sidwell, (eds.) The *Handbook of Austroasiatic Languages,* 2.1186–1226. Part II, Ch. 20. Leiden: Brill

Ring, Hiram. 2015. *A Grammar of Pnar*. PhD thesis, Singapore: National Technical University.

PRONOMINALISATION IN SOUTH ASIAN LANGUAGES: OF PEOPLE AND THEIR ACTIONS[*]

Tanmoy Bhattacharya

By looking at agreement in Indo-Aryan (IA), and Munda languages, it is suggested that the apparent similarity of a certain phenomenon across unrelated languages is not a sufficient condition for contact induced change. In particular, I will investigate the phenomenon of "multiple agreement" in detail in Munda languages to show that it is not the same as agreement in contiguous IA languages.

1 Multiple agreement in CMP languages

There exists a possible contact situation between the "Central Māgadhan Prākrit" (CMP) languages (see Bhattacharya 2016 for an explanation of the coinage there initiated for various reasons) like Maithili, Magahi, Angika, and some North Munda (Kherwarian) languages, like Santali, Mundari, Ho, etc., as these languages are spoken in the same extended geographical area. This is perhaps the reason why Chatterji(1926) obviously pointed towards these latter group of languages when faced with the phenomenon of multiple agreement (or "intricacies of the verbal system," see the quotations below) in CMP languages, although a careful reading of earlier scholars like Grierson (1903), would have revealed the verb agreement pattern for him.

It is clear from the following quotation that the multiple agreement phenomenon of the CMP languages was obvious to Grierson:

> "The principal difficulty to the beginner in the study of Maithili, is the bewildering maze of the verbal forms." … "This is due to the fact that the verb agrees not only with the subject, but with its object." (Grierson 1903: 25)

However, for Chatterji (1926), these languages held the following features:

i. "The verb-system of Maithili and Magahi seems to be a rather late development, originating or asserting itself long after the differentiation of the Māgadhī speeches."

ii. "… the *intricacies* of the later Maithili were absent in Old Maithili."

iii. Talking about the language of Vidyāpati (14th C), Chatterji states that "… especially noticeable is the *simplicity* of verb-system, with its *freedom* from the *ramifications* of pronominal infixes and affixes."

iv. Chatterji conjectures further that the pronominal affixation could be due to influx of Kōl people from South, first as "Chikā-Chikī" dialect and later spreading further. (all italics are mine)

Chatterji thus talks about a possible contact situation "from south" of the CMP language area with Munda languages synchronically co-existing; he doesn't necessarily talk about the path of Munda migration into the area.

However, there exists certain differences between the two agreement systems, that apart from highlighting the well discussed difference between agreement and cliticization, emphasizes a clear difference between the two types. Thus, these evidences point towards a direction of independent development of each type in the respective languages.

2 Empirical facts of clitics in Munda and agreement in CMP languages

The various differences between the empirical facts of the two phenomena in these two groups of languages are discussed below (from Bhattacharya, 2016).

(a) Clitics are forms of pronominals

Table (1) shows that in Mundari the clitics and the pronominals are quite similar. In the CMP languages, on the other hand, as in (1a) in Maithili, the pronominals for 'he (hon.)' and 'you (non-hon.)' are *o* and *tora*, whereas the agreement marker carrying the 3Hon+2Non-hon fused morpheme is *thunh*.

---

[*] [Keynote speech delivered at the 38th Annual Conference of Linguistic Society of Nepal]

Table 1: Mundari agreement markers

	SG	DL	PL
1st Inclusive	- ñ/ -añ	-laŋ/ -alaŋ	-bu/ -abu
1st Exclusive	---	-liŋ/ -aliŋ	-le/ -ale
2nd	-m/ -am	-ben/ -aben	-pe/ -ape
3rd	-el-il-eʔ/-lʔ/ aeʔ	--kin/ -akin	-ko/ -ako

(1) o   tora   dekh-əl-thunh
     He.H you.NH.ACC/DAT see-PAST(3H+2NH)
     'He (H) saw you(NH).'

[Yadav (1996)]

(b) Optionality of agreement marking

Although optionality of object marking is noticed in Sora (a South Munda language), in the CMP languages, such agreement cannot be optional:

(2) iɛr-ai-ɛn-a
     go/come-CLOC-N.SFX-GEN
     tiki   aninji   gudeŋ-le
     after   they   call-PST
     'After he came, he called them.'

[Anderson and Harrison, 2008:330]

(c) Bi-personal verb forms are not the norm

Either the subject (in (3a) with an intransitive verb) or both the subject and the object (as in (3b) with a causativized form of the same intransitive verb) is marked; also the subject agreement clitic (*ko* and *eʔ* in bold) is found on the Locative in (3a) and the causative subject in (3b), respectively, whereas the object clitic is incorporated into the verb in (3b) here:

(3) a. hon-ko   ote-re=**ko**   dub-ke-n-a
         child-PL ground-LOC=3PL:SUBJ sit-COMPL-INTR-IND
         'Children sat on the ground.'

     b. Sona   hon-ko=**eʔ**   dub-ke-d-**ko**-a
         Sona   child-PL=**3SG:SUBJ**   sit-COMPL-TR-**3PL:OBJ**-IND
         'Sona made the children sit.'

[Mundari, Osada 2008:121]

Although Sora is a south Munda language not ever in contact with languages in Bihar, it also has a rare instance of "bi-personal" verb form, that is,

whenever there is another argument in the sentence with a different person feature, the verb is marked for both (as in (4)). However, marking both the arguments in the verb is not the norm.

(4) əʔ-gij-lɛ-**be-ji**
     neg-see-pst-1pl-3pl
     'We didn't see them.'

[Sora, Anderson and Harrison, 2008:328-330]

Note thus that in (3b), the verbal form is marked with only argument clitic (object and subject, respectively), bi-personal verbal form of agreement, as in (4) are not the norm. Again, this feature of single argument agreement is a departure from the multiple agreement we notice in CMP languages.

(d) Pro-clitic split

The Munda languages have pro-clitics which often split across the verb and a pre-verbal word; for example, the subject clitic in the following is on the object and the object clitic is on the verb, shown here in bold (as in (3b)):

(5) a. pusi-kin seta-ko=**kin**   hua-ke-d-**ko**-a
         cat-DL   dog-PL=3DL:SUBJ   bite COMPL-TR-3PL:OBJ-IND
         'The two cats bit the dogs.'

     b. seta -kin pusi-kin=**ko**hua-ke-d-**kin**-a
         dog-PL   cat-DL=3PL:SUBJ   bite-COMPL-TR-3D:OBJ-IND
         'The dogs bit the two cats.'

[Mundari, Osada 2008:108]

The verb here carries a single clitic agreement morpheme (representing the object); this is different from CMP languages (see (c) above).Observe also the fact that the phenomenon of pro-clitic split is specific only to Munda languages with accompanying cliticization.

(e) Presence of applicative suffixes

Applicatives are employed to mark indirect objects: in Santali and Mundari, the indirect object is marked with an applicative morpheme, -*a* in the first case and –*ma* in the latter with the verb:

(6) a. dal-**a**-ŋ-a-**e**

strike-APPL-1SG:OBJ-FIN-3SG:SUBJ
'He strikes/ will strike for me.'
[Santali: Ghosh 2008: 55]

b. am   seta-ko=**ñ** om-a-**ma**-ta-n-a
2S   Gdog-PL=1SG:SUBJ give-BEN-
APPL-2SG-PROG-INTR-IND
'I am giving the dogs to you.'
[Mundari, Osada 2008:122]

Use of applicatives is not found in CMP languages.

Summarising the observation so far, let us focus on the difference that is relevant for our purpose here in this paper, namely, the order of affixes in Santali (and Munda in general):

(7)   V-Asp/Aux-**Agr$_o$**-Mood/C/Fin-**Agr$_s$**

Whereas, in CMP languages, as shown in Bhattacharya (2016), the order of affixes is:

(8)   V-Agr$_s$-Agr$_o$

Along with the clear difference between the status of agreement morphemes of the two languages; the morphemes denote clitics in Santali, and in CMP languages, they denote agreement affixes.

3 The origin of pronominalisation

For Max Müller (*The languages of the Seat of the War in the East*: with a survey of three families of language, Semitic, Arian, and Turanian. 1855. London: Williams &Norgate, p86) "(T)he third family is the Turanian. It comprises all languages spoken in Asia or Europe not included under the Arian and the Semitic families, with exception perhaps of the Chinese and its dialects." Although Hodgson had earlier in 1849 included Chinese too in this group, which hints at his later coinage of a separate family called "Tibeto-Burman": "Tamulians, Tibetans, Indo-Chinese, Chinese, Tangus, Mongols, and Turks are so many branches of another single family, viz., the Turanian" (1849 p:3). The 19th century European construct "Turanian" thus became the waste-basket category where widely unrelated languages were dumped; partly due to their overzealousness in "mapping" the world, and partly dictated by war reasons. This latter point is highlighted in Max Müller when he reasons that Tungusic languages, which extend from north China to Siberia up to the river Tunguska, are to be discounted since "they are not likely to appear on the theatre of war;" similarly, the Mongolic languages.

Though Hodgson (1849) 'unites' the Himalayas, Indo-China and Tibet as speaking languages of the same family (TB) that is marked by "syntactic poverty", among other traits, Hodgson (1856) lists a series of facts, one of them being verb pronominalisation, which, according to him, offers evidence of genetic relation between the Turanian languages; this also gave rise to the "Austric" thesis (Bhattacharya 2017). The pronominal system of the Turanian languages according to Hodgson are "greatly developed" and consist of the following traits:

a. Separate forms for personal (independent) and possessive forms of pronouns;

b. Separate inclusive and exclusive forms for 1st person pronouns;

c. Different sets of possessive pronouns: one used disjunctively (i.e. as a free form) and the other conjunctively (i.e. as an affix);

d. Distinction between dual and plural number categories;

e. Verb pronominalisation;

f. Prefixation of noun possessive forms and suffixation of verb pronominal affixes;

g. A prevailing verb structure consisting of root + transitive/intransitive marker + pronominal suffix;

h. The morphological conflation of 2nd and 3rd persons in TB and Dravidian in opposition to 1st person forms.

As we shall see many of these observations, noted more than 150 years ago, are accurate. He further considers that the Himalayish and Munda languages show pronominalisation in fullest form while other Turanian languages either lack it entirely or show much more impoverished forms. To his credit though, Hodgson does not even hint at a directional view of the spread of this feature from, what many people considered, substratal Munda to TB. However, the substratum thesis was a very popular one in the 19th century language studies in and around India, and Konow, being in-charge of parts of the *LSI* ((3(1) and

1(1)), drew a directional link between Himalayan TB languages and Munda by proposing that substratum Munda influence is the cause of pronominalisation in the former:

"It therefore seems probable that Mundas or tribes speaking a language connected with those now in use among the Mundas, have once lived in the Himalayas and have left their stamp on the dialects there spoken at the present day" (*LSI* 3(1)s179 and 1(1): 56).

Of course, there have been other theses, for example, Henderson:

"It appears not unlikely that improved knowledge of the Chin languages and of others equally remote geographically from the so-called pronominalized groups will bring further similarities to light. In this event linguists may be obliged to conclude that, contrary to what has often been supposed, pronominalisation is after all a genuine Tibeto-Burman family trait" (1957:327).

However, as Bauman (1975), convincingly argues, pronominalisation as a feature is widely distributed across North, Northwest, Northeast and Indo-China, which gives credit a native origin within TB of pronominalisation theory. As Maspero (1946) had shown, Munda and TB verb are syntactically dissimilar. For one, the pronominal affixes with the verb in TB languages seem like agreement markers and not clitics as in Munda. This has, as Bauman argues, something to do with lack of morphological case markings in Munda and their presence in TB. The disambiguation of the NOM/ ACC marked NPs in Munda is done in the verbal markings, whereas in TB, ambiguities can be recovered through case marking. There are three areas of differences between the pronominals in the two groups of languages that Bauman points out: (i) TB has more alternate pronominal forms than Munda; (ii) fixed position of affixes in TB as opposed to say in Santali where the subject clitic can be either on the preverbal element or appear verb-finally (Bodding, 1929:49). [So the word-finality cannot be of recent origin as conjectures in Hock (2013)]. Also, in Santali, and Munda in general, the clitic forms are easily derivable from the pronominal forms, but this not the case in case of TB

pronominal affixes. (iii) the Munda clitics are simpler than the TB ones, it's their order relative to the root that defines them as an object. In short, the pronominalisation system of TB languages is much more complex than the Munda languages, and in fact, Pinow (1966) concludes that:

"In proto-Munda...the pronouns properly were independent, isolatable free forms. The affix character of the pronouns, which were incorporated into the verb complex as subject or object respectively, is of more recent date" (1966:183).

The classification of pronominalized languages into subgroups at least established the fact that no one scheme fits them all, which in turn means that this feature is an archaic TB feature.

Classification of pronominalized languages modified from Bauman (1975) based on Shafer (1974)

4 Clitics versus affixes

There are various well-known diagnostics to test the differences between clitics and affixes. For Santali, these tests are indicated in Kidwai (2006), following Zwicky and Pullum (1983), and summarised below; it will be noted that Subject marker is more clitic-like than the Object marker:

a. Santali Subject clitic is not selective in terms of the category of the host, the latter may be nouns, postpositions, light verbs, adverbs, negation, etc. The object clitic however always appears after the Tense/ Aspect marker following the verb root, though it's not restricted as to the nature of the Tense/Aspect marker;

b. Unexpected gaps may appear with affixes (e.g.

stride doesn't have a past participle), but not with clitics;

c. Santali clitics do not have idiosyncratic semantics, unlike affixes (-*er* in English);

d. Clitics cannot induce stem allomorphy;

e. Clitic placement is syntactically conditioned; the distinction between the Santali finite marker –*a*, which always appears verb-finally, and the subject clitic, which can be shown to be syntactically predicted (see below);

f. Clitics attach outside of affixes; except for the finiteness marker –*a*, the subject clitic seems to be placed mostly after all the other affixes

5 Against phonologically driven clitic-placement

With regards to (v) above, note that there exists a vast literature on so-called 2nd position clitics, which have been argued to be phonologically as well as syntactically conditioned. However, this literature is mostly based on research on languages in Europe. Hock (2013) is guided by such concerns, as well as trying to look for only a historical explanation of the phenomenon. Hock (2013) reasons that just because preverbal focus in SOV languages makes the rheme stronger than the theme, it also attracts the clitic. However, this assumes two things: (i) clitic-placement is a matter of prosody, and (ii) clitic-placement is centred around finding the prosodically 'strongest' element; neither of which is a proven fact.

The hypotheses listed in Hock (2013) are also driven mostly by historical concepts. For example, his contention that Munda clitics started out as clause-initial theme position only to shift later to the Wackernagel position, and it only later moved to the postverbal position via the preverbal focus position, in short, a rightward drift theory. However, later summarising the findings, Hock (2013) contends that the rightward drift to a Wackernagel position within the domain of the rheme; note that this is really a statement about the syntactic positioning of the element in question. Further on, other alternative possibilities are considered, which include, among others, placement of the clitic (as in Serbo-Croatian) after the first element following a prosodic break after the clause-initial position. For example, in Parachi (a Southeastern Iranian language), clitics attach to

the leftmost element; however, it may also appear at the end of the verb, if there is no other constituent, or it may preferably attach to the constituent immediately preceding the verb; and alternatively attach to an earlier constituent for emphasis. This hints at too many possibilities and suffers from the same problem that many historical and typological accounts do – they never predict the non-availability of a position where a clitic can be placed.

Also if Hock's account were to be right, we would need to see data with other focus elements in the clause (even within the rheme) attracting the clitic (as the final alternative in Parachi indicated above), as far as I can tell, this does not happen. Given the main "drift' in the article, it seems that all that it is certain about is the preverbal position as far as Munda is concerned; from the mere way this is proposed, it is clear that that is a syntactic position, rather than phonology guiding clitic-placement. The example in (11d) below also clearly argues against a phonologically conditioned clitic-placement theory, where the subject clitic is placed after the negation particle. In conclusion, it may be said that clitic-placement is determined both by Syntax and Phonology

6 Unmarked positions of clitics in Santali

With regards to clitic-placement in Santali, older scholarship noticed the phenomenon carefully, although various authors described them as short forms of pronominal elements (Hoffman (1903) for Mundari, Burrow (1915) for Ho, Macphail (1953) for Santali, etc.). For example, Hoffman (1903), provides the following example:

(9) ɲel-ko-tan-a-le
     see-**them**-PRES-FIN-**we**
     'We are seeing them'

Here, the pronouns in the gloss are highlighted to indicate how Hoffman viewed those elements; thus, *ko* is the direct object and *le* is the subject for Hoffman in (9). However, Hoffman also remarks that: "These pronominal subjects when suffixed to Mundari Transitive or Intransitive Predicates give the latter a *semblance of a conjugation* ..." Recall that European scholars studying these pronominalized languages in the 19th century had very strange views about this

issue. Thus, with regards to pronominal complexity Hodgson comments that "when viewed in connection with the paucity of true conjugational forms [recalls] the fine remark that rude people think much more of the actors than the action" (1856:135).

Macphail's (1953) view on these clitics can be understood from the following:

"If the subject of the sentence, whether noun or pronoun is animate, the short form of the pronoun is always shown on the verb. Sometime it follows the verbal 'a' at the end: more often it is attached to the previous word. If the subject is an animate pronoun, it appears first in its full form and is repeated in the short form in the verb. If a noun, the corresponding pronoun is shown in the verb. If the subject is inanimate, no pronoun is shown on the verb."

To understand the unmarked position of the subject clitic, consider the following examples from Macphail (1953):

(10) a. amdɔ-m   bes-ge-a
         youEMP-2s well-EMP-FIN
         'You are well.'

     b. iɲ. dɔ   sapha-ge-ān(ān = a-iɲ)
         IEMP   clean-EMP-FIN-1s
         'I am clean.'

Other examples of subject clitic placement are as follows (from Hansdah&Murmu (2005), henceforth HM):

(11) a. iɲsen-okˀ-a-iɲ
         Igo-FUT-FIN-1s
         'I shall go.'

     b. iɲ-iɲsen-okˀ-a
         I-1sgo-FUT-FIN
         'I shall go.'

     c. iɲbazaar-iɲsen-okˀ-a
         Imarket-1sgo-FUT-FIN
         'I shall go to the market.'

     d. iɲbazaarba-iɲ   sen-okˀ-a
         Imarket-1sNEG-1sgo-FUT-FIN
         'I shall not go to the market.'

Thus, with regards to the placement of the subject clitic, we can consider the default or unmarked positions as follows:

(12) Default/ Unmarked position of the subject clitic in Santali:
         a. stem-finally (postverbally), or
         b. affixed to the preverbal element (ditropic)

Example (11a) shows the postverbal positioning, and (11b,c,d) show preverbal placement of the subject clitic *iɲ* in Santali. Although Hock (2013) reports as Osada (2008) (for Mundari) noting that the postverbal positioning is preferred by younger speakers, it must remembered even Bodding (1929:49) noted the postverbal position as a default position of the subject clitic in Santali. Example (11c) also clearly shows that clitic-placement is not a matter of phonology as it is not in the expected 2nd position of the clause.

Placement of the object clitic can be ascertained from the following (besides (9) above):

(13) a. ɲɛl-iɲ-a-e
         see-1s-FIN-3s
         'He will see me.'

     b. ɲɛl-me-a-e
         see-2s-FIN-3s
         'He will see you.'

There is also the option of fully specifying the pronouns themselves in the clause as well as the following shows:

(14) a. uniiɲɲɛl-iɲ-a-e
         heIsee-1s-FIN-3s
         'He will see me.'

     b. uniamɲɛl-me-a-e
         heyousee-2s-FIN-3s
         'He will see you.'

Thus, it is not true, as noted by both Kidwai (2005) and Hock (2013), that the subject clitic appears stem-finally only when all arguments of the verb are incorporated into the verb.

Cliticisation in Santali is, however, restricted to [+animate] nouns only, this is shown below for both subject and object clitics:

(15) a. noa   baha   dɔ   monˀ-a
         this   flower   EMP   beautiful-FIN
         'this flower is beautiful.'

   b. ɲɛl-ked-a-e
       see-PST-3s-FIN-1s
       'I saw it.'

Note that both direct and indirect object may not be cliticised, as in the following noted by Macphail (1953):

(16) a. gidradɔmæjiu-iɲem-ade-a
         childEMPwoman-1sgive-PST.3s(DAT)-FIN
         '(I) gave the child to the woman.'

   b. mæjiu-thengidra-iɲem-ked-e-a
       woman-tochild-1sgive-PST-3s(ACC)-FIN
       '(I) gave the child to the woman.'

(16) above shows that the internal argument that is not cliticised is moved out. Note also that –ade- in the (a) example is shortened form of *ked-ae* where –*ke* is dropped and –*a* (as a Dative marker, as per Hoffman (1903) or Mundari) moves to the front. This dative marker is further visible in the following from HM:

(17) a. iɲām-iɲ   ɲur-me-a
         Iyou-1s   drop-2s-FIN
         'I shall drop you.'

   b. iɲāmmidizdhiri-ɲɲur-ā-ām-a
       Iyouonestone-1sdrop-DAT-2s-FIN
       'I shall drop a stone on you.'

Note that the direct object is [-animate] in (17b) and, therefore, it does cliticise(although in English it's a PP).

Note also that the object clitics is placed after the Tense/Aspect infix, although it may not be clear form the above examples; see example below (from HM):

(18) a. ɲɛl-et-ko-kən-a-ɲ
         see-PRS-3PL-be-FIN-1S
         'I am seeing them.'

   b. ɲɛl-et-ben-kən-tahẽnkaən-a-ɲ
       see-PRS-2DL-be-PST-FIN-3SG
       'I was seeing them two.'

So, the position of the object clitic in Santali is:

(19) Default/ unmarked position of the object clitic in Santali:

Between the Tense/Aspect marker and the Finiteness marker, or:

   V-T/ASP-CL$_{OBJ}$-FIN

Like other north Munda languages like Mundari and Ho, Santali also shows the unique property of cliticising [+animate] possessive Genitives, as shown below:

(20) a. gidra   menakˀ-ko-tiɲ-a
         child   exist-3PL-1.POSS-FIN
         'I have children.'

   b. (uni)   hopon-e idi-ketˀ-e-tiɲ-a
       (he)son-3stake-PST.TRAN-3S-1.POSS-FIN
       '(He) took away my son.'

However, it is not mandatory for the genitive to cliticise on to the verb, as the following shows:

(21) a. hopon-tiɲ-e   hecˀ-en-a
         son-GEN-3S   come-PST.INTR-FIN
         'My son came.'

   b. iɲhopon-e   hecˀ-en-a
       1S-son-3S   come-PST.INTR-FIN
       'My son came.'

   c. iɲ-icˀ   hopon   hecˀ-en-a
       1S-GEN   son   come-PST.INTR-FIN
       'My son came.'

Finally, psychological predicates in Santali seem to show a phenomenon which violates the clitic-placement typology we have seen so far; this is shown below:

(22) a. rabaṅ-ketˀ-pe-a
         cold-PST-2PL-FIN
         'You were cold.'

   b. reṅgecˀ-ed-iɲ-kən-a
       hunger-PRS-1S-BE-FIN
       Lit. It hungers me ('I am hungry')

   c. tetaṅ-ko-a
       thirst-3PL-FIN
       Lit. It will make them thirsty ('They will be thirsty')

7 Analysis

To come back where we started from, namely, debating the influence of Munda languages on the CMP group of languages in terms of multiple agreement, it was pointed out that there is a crucial difference between the two groups in terms of the order of agreement morphemes or clitics; this is summarised below:

(23) V-Asp/Aux-**Agr$_o$**-Mood/C/Fin-**Agr$_s$**

(Munda)

(24) V-T-Asp-Agr$_s$-Agr$_o$ (CMP)

Thus, the relative orders of the subject and object-indexation morphemes differ in the two groups of languages.

Agreement affixes (or clitics) are assumed to be derived through the operation of "Agree"—a dependency relation between an inflectional head and the arguments in its domain, which results into the valuation of appropriate Φ-features on the head (T here). Thus, when T agrees with the Subject DP, the latter's Φ-features are copied onto T, and are then "relayed" onto the verb. This is standard Agree and is shown in (25).

(25)
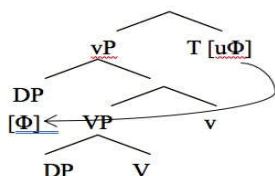
This will produce the sequence of morphemes as below:

(26)   V-T-AGR$_{SUBJ}$

However, in order to establish multi-argument agreement (as reported above for both IA and Munda), standard Agree as in (25) is not sufficient. In order to account for the fact of multi-argument agreement, the *v* head can establish Agree with another DP-argument, namely the object DP. This possibility is sketched in (27).

Note that here the Φ-features are valued on two different heads T and *v*, respectively, and two different cycles.

(27)
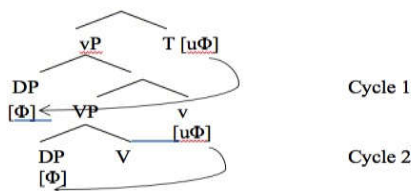
Accordingly, the "relay" of the valued features is copied further onto the [V-*v*] complex to produce the sequence as follows:

(27) V-T-AGR$_{SUBJ}$ −AGR$_{OBJ}$

However, note that this is not the order of agreement morphemes in Munda (cf. (23)). The analysis suggested for the CMP group of languages in Bhattacharya (2016) makes crucial use of the phenomenon of "Cyclic Agree" (CA) of Rezáč (2003) and Béjar (2003), which proceeds in a bottom-up fashion. Thus, in CA, the VP agreement is done first and then the derivation proceeds to do the inflectional subject agreement, by definition therefore, CA will obtain a bottom-up agreement pattern; namely, that the object agreement marker will be nearer to the verb than the subject agreement marker. This is shown in (29).

(29)
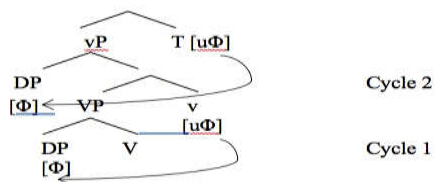
Producing a sequence such as the following, as desired for Munda languages:

(28)   V-AGR$_{OBJ}$ −T-AGR$_{SUBJ}$

There are many other details of the analysis presented as well as much of the data presented above, have all not been taken into account (though see Bhattacharya 2017), but rather the attempt has been to sketch an overall analysis based on the new notion of Cyclic Agree.

8 Conclusions

The argument presented in the first part of the paper suggests that it is possible to map a

Sprachbund in the eastern foothills of the Himalayas based on the phenomenon of multiple-agreement, that has emerged without necessarily inducing a contact situation. Based on various diagnostic tests, we have shown that multiple-agreement may involveagreement in CMP languages, whereas cliticization in Munda languages.

Furthermore, the analysis presented has shown that the syntax of multiple agreement may involve two different types of Agree relations, namely, CyclicAgree and Standard Agree, for these two different groups of languages.

References

Anderson, Gregory and K. David Harrison. 2008. Sora. In Gregory D.S. Anderson (ed.), *TheMunda Languages.* London / New York: Routledge. [Routledge Language Family Series], 299-380.

Bauman, James J. 1975. Pronouns and Pronominal Morphology in Tibeto-Burman. University of California, Berkeley, Ph.D.

Bhattacharya, Tanmoy. 2018. The syntax of argumentindexation in the languages ofIndia: A case of Macro- or Microvariation? Paper presented in the Workshop on Variation in Languages, CEL, Sikkim University [March].

Bhattacharya, Tanmoy. 2017. Peopling of the Northeast:Part 3. *neScholar*, vol 3:1, 61-70.

Bhattacharya, Tanmoy. 2016. Inner/Outer Politeness in Central MāgadhanPrākrit Languages: Agree as Labeling. *Linguistic Analysis*, vol 40, 3-4, 297-336.

Chaterji, Suniti Kumar. 1926. *The Origin and Development of the Bengali Language.* GeorgeAllen & Unwin, London.

Grierson, G. A. 1887. *Seven Grammars of the Dialects and Subdialects of the Bihari Language: Spoken in the Province of Bihar, in the Eastern Portion of the North-Western Provinces, and in the Northern Portion of the Central Provinces. Part VII: South Maithilī-Bangālī dialect of south Bhagalpūr*. Printed at the Bengal Secretariat Press: Calcutta.

Grierson, G. A. 1903. *Linguistic Survey of India*, Vol V, Indo-Aryan Family, Eastern Group,Part II: Specimen of Bihārī and Oriya Languages. Office of the Superintendent, Government Printing, Calcutta, INDIA.

Grierson, G. A. 1909. *Linguistic Survey of India*, Vol III, A Manual of the Kashmiri language. Office of the Superintendent, Government Printing, Calcutta, INDIA.

Ghosh, Arun. 2008. Santali. In Gregory D.S. Anderson (ed.), *The Munda Languages.* London / New York:Routledge. [Routledge Language Family Series]: 11-98.

Hansdah, R.C. and Murmu, N.C. 2005. *A Framework for Learning and Understanding Santali in OlChiki Script.* Wesanthals e-group Technical Report No. 1.

Henderson, Eugenie J. A. 1957. Colloquial Chin as pronominalized language. 3*SOAS* 20:323-27.

Hock, Hans Henrich. 2013. Backernagel is Wackernagel lite. On the "P-minus 2" clitics of Santali. *Lingua Posnaniensis*, vol. LV (2). The Poznań Society for theAdvancement of the Arts and Sciences.

Hodgson, Brian. 1849. [I849d] On the aborigines of north-eastern India. *JASB* 18:350-59. [reprinted in Miscellaneous essays relating to Indian subjects, Vol. 2, 1— 10. 1880].

Hodgson, Brian. 1856. Aborigines of the Nilgiris with remarks on their affinities. JASB 25:498-522. [reprinted in Miscellaneous essays relating to Indian subjects, Vol. 2, 125-44. 188Q].

Hoffman, Rev. J. 1950. *Encyclopedia Mundarica*, 13 Vols, Patna: Superintendent of Government Printing, Bihar.

Kidwai, Ayesha. 2005. Santali 'Backernagel' Clitics: Distributing Clitic Doubling." In: Bhattacharya, T.(ed.) 2005. *The Yearbook of South Asian Language and Linguistics*; New York, Mouton: 189–207.

Macphail, R. M. 1953. *An Introduction to Santali.* Calcutta: Firma KLM Private limited.

Müller, Max. 1854. Le.er to Chevalier Bunsen on the Classification of the Turanian Languages.

Müller, Max. 1855. *The languages of the Seat of the War in the East*: with a survey of three families of language, Semitic, Arian, and Turanian. London: Williams &Norgate.

Osada, Toshiki. 2008. Mundari. In Gregory D.S. Anderson (ed.), *The Munda Languages.* London / New York: Routledge. [Routledge Language Family Series]: 99-164.

Yadav, Ramawater. 1996. *A Reference Grammar of Maithili.* The Hague: Mouton.

LIST OF LIFE MEMBERS OF LINGUISTIC SOCIETY OF NEPAL

	No.	Name
1	001	Ramawatar Yadav, <ramyadav992@live.com>
2	002	Kamal Prakash Malla Maitidevi, <mallacorp@gmail.com>
3	003	Chandra Devi Shakya, Lalitpur.
4	004	Yogendra P. Yadava, CDL, Kirtipur, <ypyadava@gmail.com>
5	005	M.S. Ningomba, Manipur University, India.
6	006	Bernhard Kolver, University of Kiel, Germany.
7	007	Ulrike Kolver, Germany.
8	008	Shailendra Kumar Singh, P. K. Campus, Kathmandu.
9	009	Burkhard Schottendreyer, Spartado Aereo 100388, Columbia.
10	010	Tika Bahadur Karki, American Peace Corps, Kathmandu (late)
11	011	Richard R. Smith, United Mission to Nepal, Kathmandu.
12	012	Horst Brinkhaus, University of Kiel, Germany.
13	013	John P. Ritchott, USA.
14	014	Subhadra Subba, Kirtipur.
15	015	Ross C. Caughly, Australia, <ross_caughley@sil.org >.
16	016	James J. Donelly, St. Xavier's School, Kathmandu.
17	017	Nishi Yoshio, University of Kyoto, Japan.
18	018	Shreedhar P. Lohani, Maitidevi, Kathmandu.
19	019	Tika P. Sharma, Mahendra Ratna Campus, Tahachal, Kathmandu.
20	020	Ronald Bielmeier, University of Berne, Switzerland.
21	021	Ian Alsop, Panipokhari, Kathmandu.
22	022	Ballabh Mani Dahal, Central Department of Nepali, Kirtipur. (Late)
23	023	Colin S. Barron, Britain
24	024	Shishir Kumar Sthapit, Baneshwar.
25	025	Chuda Mani Bandhu, CDL, Kirtipur, <cmbandhu@yahoo.com>.
26	026	Tej Ratna Kansakar, CDL, Kirtipur, <t.r.kansakar@gmail.com>.
27	027	Rameshwar P. Adhikari, Kathmandu.
28	028	Nirmal Man Tuladhar, LinSuN, CDL, Kirtipur, <nirmal_tuladhar2006@yahoo.com>.
29	029	Abhi Narayan Subedi <abhi@mail.com>.
30	030	Beverly Hartford, Indiana University, USA.
31	031	J. P. Cross, Pokhara.
32	032	Marshal Lewis, Indiana University, USA.
33	033	K.V. Subbarao, CDL, Delhi University, India.
34	034	Devi P. Gautam, CDN, Kirtipur.
35	035	Chandra Prakash Sharma, Kathmandu.

36	036	Werner Winter, Germany.
37	037	Baidyanath Jha, R. R. Multiple Campus, Janakpur.
38	038	Satya Ranjan Banerjee, Calcutta University, India.
39	039	George van Driem, G. P.O. Box 991, Kathmandu, <george.van.driem@gmail.com>.
40	040	Beerendra Pandey, CDE, Kirtipur.
41	041	Kalpana Pandey, c/o Beerendra Pandey.
42	042	Chandreshwar Mishra, DEE, Kirtipur.
43	043	Rudra Laxmi Shrestha, Lalitpur.
44	044	Khagendra K.C., Patan Multiple Campus, Patan.
45	045	Shambhu Acharya, Patan Multiple Campus, Patan. (late)
46	046	Durga P. Bhandari, Baneshwar, Kathmandu.
47	047	Rajendra P. Bhandari, CDE, Kirtipur.
48	048	Shree Krishna Yadav, Kathmandu.
49	049	Jai Raj Awasthi, DEE, Kirtipur, <jrawasthi@gmail.com>.
50	050	Novel K. Rai, LinSuN, CDL, Kirtipur, <drnovel4@yahoo.com>.
51	051	Padma P. Devakota, Maitidevi, Kathmandu.
52	052	Manfred G. Treu, CIL, Kathmandu.
53	053	Gautami Sharma, P. K. Campus, Kathmandu.
54	054	Sangita Rayamajhi, CDE, Kirtipur.
55	055	Bhuvan Dhungana, R. R. Campus, Kathmandu.
56	056	Baidaya Nath Mishra, Trichandra Campus, Kathmandu.
57	057	Krishna Pradhan, Saraswati Campus, Kathmandu.
58	058	K.B. Maharjan, Bernhardt College, Balkhu.
59	059	Hriseekesh Upadhyay, R. R. Campus, Kathmandu.
60	060	Nayan Tara Amatya, Kathmandu.
61	061	Sueyoshi Toba, G. P.O. Box 991, Kathmandu, <si_toba@sil.org>.
62	062	Sanjeev K. Uprety, CDE, Kirtipur.
63	063	Anand P. Shrestha, Central Department of English, Kirtipur.
64	064	Bishnu Raj Pandey, Kathmandu.
65	065	Mohan Prasad Banskota, Saraswati Campus, Kathmandu. (late)
66	066	Rupa Joshi, P. K. Campus, Kathmandu.
67	067	Keshab Gautam, Saraswati Campus, Kathmandu.
68	068	Krishna Kumar Basnet, Saraswati Campus, Kathmandu.
69	069	Nirmala Regmi, P. K. Campus, Kathmandu.
70	070	Jyoti Tuladhar, Kathmandu.
71	071	Bidya Ratna Bajracharya, Samakhusi, Kathmandu.
72	072	Shanti Basnyat, Kathmandu.
73	073	Madhav P. Pokharel, CDL, Kirtipur, <madappokhrel@yahoo.com>.

74	074	Ramchandra Lamsal, Department of Nepali Education, Kirtipur.
75	075	Arun Kumar Prasad, Trichandra Campus, Kathmandu.
76	076	Bijay Kumar Rauniyar, CDE, Kirtipur, <bijayrauniyar@hotmail.com>.
77	077	Sajag S. Rana, Patan M Campus, Patan, Lalitpur.
78	078	Balaram Aryal, CIL, Kathmandu.
79	079	Bed P. Giri, USA.
80	080	Amma Raj Joshi, CDE, Kirtipur.
81	081	Parshuram Paudyal, CIL, Kathmandu.
82	082	Martin W. Gaenszle, University of Heidelberg, South Asia Institute.
83	083	Maya Devi Manandhar, Saraswati Campus, Kathmandu.
84	084	Pradeep M. Tuladhar, Patan Campus, Patan.
85	085	Megha Raj Sharma, CIL, Kathmandu.
86	086	Mohan Sitaula, CIL, Kathmandu.
87	087	Anand Sharma, R. R. Campus, Kathmandu.
88	088	Nanda Kishor Sinha, Birgunj.
89	089	Ram Bikram Sijapati, Patan Multple Campus, Patan.
90	090	Punya Prasad Acharya, Kanya Campus, Kathmandu.
91	091	Narayan P. Gautam, CDN, Krtipur.
92	092	Mahendra Jib Tuladhar, Saraswati Campus, Kathmandu.
93	093	Govinda Raj Bhattarai, DEE, Kirtipur, <tu.govinda@gmail.com>.
94	094	Balthasar Bickel, Max-Planck Institute, The Netherlands, <bickel@rz.uni-leipzig.de>.
95	095	Bhusan Prasad Shrestha, Saraswati Campus, Kathmandu.
96	096	Anuradha Sudharsan, India.
97	097	Bert van den Hock, University of Leiden, The Netherlands.
98	098	Harihar Raj Joshi, G. P. O. Box 2531, Kathmandu. (late)
99	099	Pramila Rai, P. K. Campus, Kathmandu.
100	100	Ram Ashish Giri, Melborne, Australia.
101	101	Tulsi P. Bhattarai, Kathmandu.
102	102	Tika P. Uprety, Devkota Memorial School, Biratnagar-2.
103	103	Mohan Himanshu Thapa, Balkhu, Kirtipur.
104	104	Anusuya Manandhar, R. R. Campus, Kathmandu.
105	105	Austin Hale, Erli-Huebli, 8636 Wald (ZH), Switzerland, <austin_hale@sall.com>.
106	106	Larry L. Seaward, USA.
107	107	Sashidhar Khanal, Trichandra Campus, Kathmandu, <sashidhar7@hotmail.com>.
108	108	Boyd Michailovsky, France, <boydm@vjf.cnrs.fr>.
109	109	Mazaudon Martine, LACITO/CNRS, France.
110	110	Tsetan Chonjore, University of Wisconsin, USA.
111	111	Sushma Regmi, Ratna Rajyalaxmi Campus, Kathmandu. (Late)

112	112	Swayam Prakash Sharma, Mahendra Campus, Dharan.
113	113	Anjana Bhattarai, DEE Kirtipur.
114	114	Nivedita Mishra, P. K. Campus, Kathmandu.
115	115	Gunjeshwari Basyal, Palpa Campus, Tansen.
116	116	Philip Pierce, Nepal Research Center, Kathmandu.
117	117	Binay Jha, R. R. Campus, Kathmandu.
118	118	Mukunda Raj Pathik, CIL, Kathmandu.
119	119	Laxmi Sharma, P. K. Campus, Kathmandu.
120	120	Bhesha Raj Shiwakoti, P. K. Campus, Kathmandu.
121	121	Siddhi Laxmi Baidya, CIL, Kathmandu.
122	122	Sulochana Dhital, CIL, Kathmandu.
123	123	Simon Gautam, Bhaktapur Campus, Bhaktapur.
124	124	Geeta Khadka, CDE, Kirtipur.
125	125	Phanindra Upadhyay, P. K. Campus, Kathmandu.
126	126	Sundar K. Joshi, Patan Multiple Campus, Patan.
127	127	Shushila Singh, P. K. Campus, Kathmandu.
128	128	Ganga Ram Gautam, Mahendra Ratna Campus, Tahachal.
129	129	Lekhnath Sharma Pathak, CDL, Kirtipur, <lekhnathspathak@gmail.com>.
130	130	Rishi Rijal, Mahendra R. Campus, Tahachal.
131	131	Binod Luitel, Mahendra Ratna Campus, Tahachal.
132	132	Anju Giri, DEE Kirtipur.
133	133	Meena Karki, Mahendra Ratna Campus, Tahachal.
134	134	Netra Sharma, Mahendra R. Campus, Tahachal, Kathmandu.
135	135	Tanka Prasad Neupane, Mahendra Campus, Dharan.
136	136	Ram Raj Lohani, CDL, Kirtipur, <ramlohani@gmail.com>.
137	137	Balaram Prasain, CDL, Kirtipur, <prasain2003@yahoo.com>.
138	138	Kamal Poudel, CDL, Bharatpur-9, Chitwan, <kpoudelnp@yahoo.com>.
139	139	Bhim Lal Gautam, CDL, Kirtipur, <bgautam2003@yahoo.com>.
140	140	Baburam Adhikari, Manamohan Memorial College, Kathmandu.
141	141	Ram Hari Lamsal, Saraswati Multiple Campus, Kathmandu.
142	142	Lekhnath Khanal, Hillbird English School, Chitwan.
143	143	Silu Ghimire, P. K. Campus, Kathmandu.
144	144	Krishna Prasad Chalise, CDL, Kirtipur, <krishna40e@yahoo.com>.
145	145	Uttam Bajgain, Baneswor, Kathmandu, <upb2062@yahoo.com>.
146	146	Bharat Raj Neupane, Nagarjun Academy, Kathmandu.
147	147	Dan Raj Regmi, CDL, Kirtipur, <danrajregmi8@gmail.com>.
148	148	Bhim Narayan Regmi, CDL, Kathmandu, <bhim_regmi@yahoo.com>.
149	149	Ajit Man Tamang, Kathmandu, <ajitman4@yahoo.com>.

150	150	Meera Shrestha, Mahendra Ratna Campus, Tahachal.
151	151	Puspa Karna, Mahendra R. Campus, Tahachal, Kathmandu.
152	152	Hemanga Raj Adhikari, Department of Nepali Education, Kirtipur.
153	153	Usha Adhikari, IOE, Pulchok.
154	154	Dev Narayan Yadav, Patan Multiple Campus. <devnarayan_yadav@yahoo.com>.
155	155	Khagendra Prasad Luitel, CDN, Kirtipur.
156	156	Rasmi Adhikari, Manmohan Memorial College, Kthmandu.
157	157	Toya Nath Bhattarai
158	158	B. K. Rana
159	159	Nabin K. Singh, CDC, Sanothimi.
160	160	Til Bikram Nembang (Bairagi Kainla), Nepal Academy, Kathmandu.
161	161	Khadka K. C., Sanothimi Campus, Bhaktapur. (Late)
162	162	Kedar Prasad Poudel, Mahendra Campus, Dharan, <kedarpdpoudel@yahoo.com>.
163	163	Ram Prasad Dahal, Modern Kanya Campus, Kathmandu.
164	164	Govinda Bahadur Tumbahang, CNAS, Kirtipur, <govindatumbahang@yahoo.com>.
165	165	Keshar Mohan Bhattarai, Chabahil, Kathmandu.
166	166	Omkareswor Shrestha, Central Department of Nepalbhasa, Patan, <omkareshwors@yahoo.com.in >.
167	167	Binda Lal Shah, Sanothimi Campus, Bhaktapur.
168	168	Deepak Kumar Adhikari, Asam, India.
169	169	Begendra Subba, Dharampur, Jhapa, <begendra@yahoo.com>.
170	170	Ganga Ram Pant, CDL, Kirtipur.
171	171	Tika Ram Paudel, Kathmandu <trpoudel@hotmail.com>
172	172	Bharat Kumar Bhattarai, CDN, Kirtipur.
173	173	Stephen Watters, Texas, USA, <steve_watters@sil.org>.
174	174	Bhabendra Bhandari, Madhumalla, Morang, <bhandaribb@yahoo.com>.
175	175	Vishnu P. Singh Rai, DEE, Kirtipur, <vpsrai@yahoo.com>.
176	176	Viswanath Bhandari, Patan Multiple Campus, Patan.
177	177	Dilli Ram Bhandari, Surkhet.
178	178	Ananta Lal Bhandari, Gulmi.
179	179	Tek Mani Karki, Kathmandu, Tahachal, Kathmandu.
180	180	Dip Karki, Mahendra Ratna Campus, Tahachal.
181	181	Nandish Adhikari, CDL, Kirtipur.
182	182	Hari Prasad Kafle, Mahendra Ratna Campus, Tahachal.
183	183	Dubi Nanda Dhakal, CDL, Kirtipur, <dndhakal@yahoo.com>.
184	184	Narayan P. Sharma, SOAS, UK, <narayan.sharma57@gmail.com>.
185	185	Krishna P. Parajuli, CDL, Kirtipur, <parajuli2025@yahoo.com>.
186	186	Nabin Khadka, Arun Valley Cultural Group, Kathmandu.

187	187	Ambika Regmi, LinSuN, CDL, Kirtipur, <ambikaregmi@gmail.com>.
188	188	Bag Devi Rai, Anamnagar, Kathmandu, <pakucha2002@yahoo.com>.
189	189	Jyoti Pradhan, CDL, Kirtipur, <jyoti100np@yahoo.com>.
190	190	Amrit Hyongen Tamang, Baneshwor, Kathmandu, <amrityonjant@yahoo.com>.
191	191	Dilendra Kumar Subba, Patan, Lalitpur.
192	192	Laxman Ghimire, Greenfield National College, Kathmandu.
193	193	Kamal Thoklihang, R. R. Campus, Kathmandu.
194	194	Lal-Shyākarelu Rapacha, Research Institute of Kiratology, Kathmandu, <kiranti.muru@hotmail.com>.
195	195	Lok Bahadur Chhetri, Ministry of Foreign Affairs, Kathmandu.
196	196	Krishna Chandra Sharma, CDE, Kirtipur.
197	197	Devi Prasad Sharma Gautam, CDE, Kirtipur.
198	198	Bal Gopal Shrestha, CNAS, Kirtipur.
199	199	Netra Prasad Paudyal, Germany, <netrapaudyal@yahoo.com>.
200	200	Rajendra Thokar, Chandranigahapur-1, Rautahat, <raaj_062@yahoo.com>.
201	201	Sulochana Sapkota (Bhusal), LinSuN, CDL, Kirtipur, <sapkota_sulochana@hotmail.com>.
202	202	Kishor Rai, GPOB 10866, Kathmandu, <raikishor@hotmail.com>.
203	203	Suren Sapkota, LinSuN, CDL, Kirtipur.
204	204	Gopal Thakur, Kachornia-1, Bara, <gtbs@ntc.net.np>.
205	205	Dhruba Ghimire, Saraswati Campus, GPOB 13914, Kathmandu, <sugdhi@enet.com.np>.
206	206	Laxmi Prasad Khatiwada, Kathmandu, <lkhatiwada@mpp.org.np>.
207	207	Kamal Mani Dixit, Madan Puraskar Pustakalaya, Patan Dhoka, <kmldxt@wlink.com.np>.
208	208	Gelu Sherpa, Baudha, Kathmandu, <g.sherpa@gmail.com>.
209	209	Bal Mukunda Bhandari, Department of English Education, Kirtipur, <bhandaribm@yahoo.com>.
210	210	Bir Bahadur Khadka, Maharanijhoda-7, Jhapa.
211	211	Chandra Khadka, Madan Bhandari Memorial College, Kathmandu, ckhadka77@gmail.com>.
212	212	Goma Banjade, Kathmandu, <gbanjade@yahoo.com>.
213	213	Ichha Purna Rai, Dhankuta Campus, Dhankuta, <timma.rai@gmail.com>.
214	214	Karnakhar Khatiwada, CDL, Kirtipur, <karnakhatiwada@yahoo.com>.
215	215	Kedar Bilash Nagila, Kathmandu <kedarnagila@yahoo.com>.
216	216	Krishna Kumar Sah, LinSuN, CDL, Kirtipur.
217	217	Krishna Paudel, Amalachaur-8, Baglung, <kpbinnyas_23@yahoo.com>.
218	218	Peshal Bastola, Canada.
219	219	Ramesh Khatri, LinSuN, CDL, Kirtipur.

220	220	Ramhari Timalsina, Kathmandu.
221	221	Prem B. Phyak, DEE, Kirtipur, pphyak@gmail.com.
222	222	Ram Nath Khanal, IOE, Pulchowk, Kathmandu.
223	223	Sajan Kumar Karna, Thakur Ram Multiple Campus.
224	224	Carl Grove, SIL International.
225	225	Indresh Thakur, Central Department of Linguistics, TU.
226	226	Jivendra Deo Giri, CDN, TU.
227	227	Kwang-Ju Cho, <kwangjucho@hotmail.com>.
228	228	Laxmi B. Maharjan, Central Department of English Education, TU.
229	229	Mahesh Kumar Chaudhari, Mahendra B. M. Campus, Rajbiraj.
230	230	N. Pramodini, University of Manipur, India.
231	231	Shova K. Mahato, LinSuN, CDL, Kirtipur.
232	232	Shova Lal Yadav, <relative_nepal@yahoo.com>.
233	233	Toya Nath Bhatta, Maryland College, Kathmandu.
234	234	Uma Shrestha, Western Oregon University, USA.
235	235	Binu Mathema, Cosmopolitan College, Kathmandu.
236	236	Dipti Phukan Patgiri, University of Guhati. India.
237	237	Diwakar Upadhyay, Manmaiju-7, Katmandú, diwakarupadhyay@hotmail.com.
238	238	Krishna Prasad Paudyal, Ratnanagar, Chitwan, <krishnapdyl@gmail.com>.
239	239	Om Prakash Singh, Prajatantra Secondary School, Morang, <omsingh33@gmail.com>.
240	240	Bishnu Chitrakar, Dallu Awas, Swayambhu, <bishnuchitrakar@hotmail.com>
241	241	Durga Dahal, Mahendra Ratna Campus, Tahachal, Kathmandu.
242	243	Durgesh Bhattarai, Madan Bhandari Memorial College, Kathmandu.
243	243	Kamal Neupane, Madan Bhandari Memorial College, Kathmandu.
244	244	Kamal Prasad Sapkota, Madan Bhandari Memorial College, Kathmandu.
245	245	Khila Sharma Pokharel, Bhairawashram, MSS, Gorkha. <neltakhila@yahoo.com>.
246	246	Niruja Phuyal, Madan Bhandari Memorial College, Kathmandu.
247	247	Purna Chandra Bhusal, Madan Bhandari Memorial College, Kathmandu.
248	248	Shobakar Bhandari, Madan Bhandari Memorial College, Kathmandu.
249	249	Bharat Kumar Ghimire, nesoa_bharat@hotmail.com.
250	250	Binod Dahal, <sabdha.binod@gmail.com>
251	251	Hemanta Raj Dahal, Chabahil, Kathmandu. <hemant_dahal@hotmail.com>
252	252	Keshab Prasad Joshi, Arunima Educational Foundation, <aefjoshi@gmail.com>
253	253	Moti Kala Subba (Deban), <motikala_d1@hotmail.com>
254	254	Rishi Ram Adhikari
255	255	Ramesh Kumar Limbu
256	256	Pam Bahadur Gurung, CDE, Kirtipur. <pam.gurung@gmail.com>
257	257	Rakesh Kumar Yadav, Imadol-8, Lalitpur. <rakeshsarlahi@gmail.com>

258	258	Tara Mani Rai, <raitaramani@yahoo.com>
259	259	Pankaj Dwivedi, India, <pankaj.linguistics@gmail.com>
260	260	Prem Prasad Poudel, Mahendra Ratna Campus, <premprasadpoudel@hotmail.com>
261	261	Ekku Maya Pun, Kathmandu University
262	262	Fatik Thapa
263	263	Vikram Mani Tripathi
264	264	Hansawati Kurmi
265	265	Padam Rai
266	266	Ganesh Kumar Rai
267	267	Amar Tumyahang
268	268	Balram Adhikari, Mahendra Ratna Campus
269	269	Upendra Ghimire
270	270	Sanjog Lapha
271	271	Dhan Prasad Subedi <subedi.dhanp@gmail.com>
272	272	Gulab Chand, India <gulab.chand@iitrpr.ac.in>
273	273	Mahboob Zahid, AMU, India, <zahidmahboob4u@gmail.com>
274	274	Mani Raj Gurung, Cosmopolitan College, Chabahil
275	275	Tek Mani Karki, Mahendra Ratna Campus, Tahachal, <tmkarki@gmail.com>
276	276	Arun Nepal, Ilam Mutiple Campus, <nplarun@gmail.com>
277	277	Netra Mani Rai, <netramrai@gmail.com>
278	278	Dr. Tanmoy Bhattacharya, Delhi University
279	279	Dr. Samar Sinha, Sikkim University
280	280	Winchamdinbo Mataina, Sikkim University
281	281	Dr. Chuda Mani Pandey, Campus of International Languages
282	282	Jagadish Poudel, DEE
283	283	Pratigya Regmi
284	284	Bhoj Raj Gautam, Teaching Hospital, Tribhuvan University
285	285	Dr. Laxmi Raj Pandit, Nepal Sanskrit University
286	286	Dr. Harish Chandra Bhatta
287	287	Shankar Subedi, Central Department of English, Tribhuvan University

Abbreviations used in this list:

CDC	Curriculum Development Centre	CDE	Central Department of English
CDL	Central Department of Linguistics	CDN	Central Department of Nepali
CIL	Campus of International Languages	CNAS	Centre for Nepal and Asian Studies
DEE	Department of English Education	IOE	Institute of Engineering