

# USE OF SOFTWARE SUPPORT IN NEPAL'S LANGUAGES

**Yogendra P Yadava**  
**Professor and Head, Central Department of Linguistics**  
**Tribhuvan University, Kirtipur, Kathmandu, Nepal**  
**yadavayp@hotmail.com**

## The Context

The initial use of computer software for Nepal's languages can be traced back to 70s when SIL linguists, in collaboration with some Nepalese linguists, collected texts from various languages including Nepali (Bandhu, 1971), created their databases, wordlists and concordances, and tried to do their linguistic analysis within the framework of tagmemic grammar. Most of these works appear in printed form or mimeograph. Later Acharya (1991), in his doctoral dissertation, used a computer programme to create an analyzed corpus of a selected Nepali written text and prepare Nepali grammar on its basis.

In 1995 Royal Nepal Academy undertook the task of compiling Prajna Nepali Dictionary, an etymological dictionary based on historical principles. To initiate this project a seminar-cum-workshop was organized to discuss issues relevant to lexicography. This seminar focused on the use of computer programmes, esp. those employed in preparing Oxford English Dictionary. Due to the lack of appropriate expertise it was then not feasible to go ahead with the use of electronic gadgets. In the meantime the Academy managed to avail the assistance of Dr Warren W Glover, who helped the project to use the computer programme, called Shoebox, in preparing a partial Nepali text database based on School-level Nepali textbooks and Gorkhapatra Nepali daily newspaper, interlinearizing it and making lexical entries. However, this dictionary project could not make much progress due to the lack of financial support.

Computer software also supported the compilation of a dictionary of Classical Newari: compiled from manuscript sources (2001), esp. for the purposes of scanning and formatting. At present, Hale has been using a computer programme for preparing Newari lexicon and grammar. Gurung-Nepali-English Dictionary (2003), edited by Warren W Glover and Ratna Bahadur Gurung, is a trilingual dictionary prepared through the application of Shoebox. Besides the Gurung lexicon, this dictionary contains two indices: Nepali-Gurung Index and English-Gurung Index. These indices have been created through the reversal process which has reversed Gurung lexicon so that it has been sorted by the Nepali and English glosses. Other recent ventures include the electronic project for documenting languages Nepal's languages at CNRS (Paris), endangered language project (as a part of Linguistic Survey of Nepal Project at Central Department of Linguistics, TU) undertaken by a joint team of Nepalese and German linguists including Balthasar Bickel, and the updating of Devanagari script by Unlimited Software and the development of Devanagari Unicode by Madan Purskar Guthi in Kathmandu.

Central Department of Linguistics, TU has recently taken interest in using electronic tools for Nepal's languages. The Languages of Nepal Information Management System (LONIMS) project, introduced by Daniel McCloy in 1997 and later pursued and completed by Peter Oatley, aims to produce a computer database structure for storing information about the languages of Nepal. It is addressed to two main types of data: the language variety data and the resource data. This project consists of four components:

- (i) the LONIMS database ( an MS Access Database holding the language and bibliographical data,

- (ii) System Administrator's Guide (a manual to the system administrator to understand the structure and design of the database,
- (iii) User Guide ( a manual to help users see how to use LONIMS) and
- (iv) Online Help File (information on how to use the database (Oatley, 2003: 48-53)

In last October, Central Department of Linguistics organized a week-long seminar on Toolbox (the latest version of Shoebox) with presentation by Dr Allen Buseman. This user-friendly programme serves as a field linguist's tool for preparing the database, lexicon and grammar of a given language.

It has been a matter of coincidence that this seminar has been immediately followed by a small grant from the National Foundation for Development of Indigenous Nationalities (NFDIN) for preparing the basic dictionaries of the five languages, viz. Baramu, Lohorung, Mewahang, Sherpa, and Gurung and documentation of the three lesser known languages, viz. Bhujel, Chantyal and Hayu. The dictionary project aims to collect a sample of natural texts for each of the languages for word elicitation, structure a lexical database and organize the collected words into lexical entries. The goals of the documentation project are to collect folklores and other cultural texts, interlinearize them with the help of text-based lexicon, and include other grammatically relevant materials to prepare the sketch grammars of the related languages and their basic wordlists.

Realizing the efficacy of software support for languages, Central Department of Linguistics has decided to introduce a basic course in computational linguistics to introduce students to basic computer skills and teach them how to use specialized software for linguistic analysis. More specifically, this course deals with the tools for dictionary making and text interlinearizing and sound and speech analysis.

## Suggestions

Software support for languages has been used for quite some time in Nepal. It has, however, been sporadic and restricted to foreign linguists and areas. In such a context there is a need for a concerted effort of linguists and computer specialists to thrash out a course of action. Some of the vital linguistic issues where computer can be of immense help may be stated as follows:

### (i) Development of expertise

Before we embark of applying software support for languages, it is necessary to develop expertise in the field of computational linguistics. It may be suggested that the M.A. course in linguistics be updated to familiarize with the latest software tools for languages.

### (ii) LONIMS database

Now that LONIMS project is complete and all its products are available on a CD-ROM and manuals, there is a need to enter data into it and upgrade the database regularly. This database will provide the existing situation about Nepal's languages including information about their identification and diverse studies carried out by various individuals and organizations in different places.

### (iii) Documentation

It is high time that the endangered languages of Nepal were documented immediately. According to the 2001 Census, there are 92 (and a few unidentified) languages spoken as mother tongues in Nepal. A quite few of Nepal's minor languages are on the verge of extinction. These languages include Kumal, Majhi, and Bote of Indo-Aryan family and Baramu, Dura, Pahari, Raute, Raji, Hayu, Mewahang, Koi, and Tilung of Tibeto-Burman family (Toba and Rai, 2002). Kusunda has 87 speakers and is also endangered though it was reported be dead earlier. Computer

programmes can be of tremendous help for field linguists in documenting such languages. They can be used to collect folktales and other cultural texts as well as discourse materials, create phonological, lexical and grammatical databases, analyze them through speech analyzer and natural language processing.

**(iv) Other applications**

- a. Educational for All (for preparing basic dictionaries and reading materials)
- b. Machine translation ( useful in a multilingual setting in Nepal)
- c. Reading machines (for blind children)
- d. Developing scripts to suit the sound system of Nepal's languages.

To sum up, these and other appropriate strategies can help the preservation and promotion of Nepal's linguistic diversity.

## **References:**

- Acharya, J.R.1991.A descriptive grammar of Nepali and an analyzed corpus, Washington, D.C.: Georgetown University Press.
- Bandhu, C.M.1971. 'The computer concordance of spoken Nepali', Okhalohoma:SIL.CBS.2001. Population Report: Kathmandu, CBS.
- Glover, W. W. and R. B. Gurung.2003. Gurung-Nepali-English Dictionary, Kathmandu: Tamu Buddha Sewa Samiti.
- Malla, K.P.ed.2001. A dictionary of Classical Newari: compiled from manuscript sources, Kathmandu: Nepal Bhasa Dictionary Committee.
- Oatley, Peter.2003. 'Final report on the research and development of a computer database system for the Central Department of Linguistics', Nepalese Linguistics, 20: 48-53.
- Toba, S., I. Toba and N.K.Rai.2002. 'Languages of Nepal:UNESCO Kathmandu Report', Kathmandu.