

# Challenges of Sherpa Dictionary Project

**Sang Yong Lee**

Korea Research Institute for Languages and Cultures

[Sang-yong\\_lee@sall.com](mailto:Sang-yong_lee@sall.com)

*A paper presented at The Asia Lexicography Conference,  
Chiangmai, Thailand  
24th – 26th May, 2004*

---

*Sherpa, which is a Tibeto-Burman language of Nepal, does not have standardization yet in dialect, script, and orthography. Involvement of the language group is important in the very first processes of dictionary production. Without consensus from the whole language group it will be difficult to get approval from the main society for the completed dictionary.*

**Dialect:** *In sociolinguistic survey to find out the more prestigious dialect among three different Sherpa dialects, the southern one was chosen.*

**Script:** *Even though a field test showed that 88% of the people wanted Devanagari script, and that script was developed and used for several years, the Sherpa society recently decided to use Tibetan script for their language in their primary school textbook. So, painful change was accepted from Devanagari to Tibetan script. This change will have effect not only in the script of the main entry, but also in sorting order in dictionary. This incident will be a worthwhile lesson for field lexicographers in similar situations to note.*

**Orthography:** *Stimulating the language group to see the need for standardization of their language, and to make some appropriate actions, is very crucial in dictionary making process. In this sense dictionary production is a basic groundwork for literacy and translation of written documents in underdeveloped language groups.*

## Contents

1	Introduction.....	2
2	The background of Sherpa language.....	3
3	Dialect decision.....	3
4	Script survey .....	4
5	The practicality and Sherpa identity .....	5
5.1	Challenges of Sherpa Dictionary project .....	5
5.2	The practicality and Sherpa identity .....	6
5.3	Problem solving .....	7
5.3.1	Tibetan sorting .....	7
5.3.2	Main entry in Devanagari with Tibetan transliteration.....	7
5.3.3	Index of Tibetan script in Appendix .....	7
6	Standardization of the language.....	7
6.1	Who does make the standardization of a language? .....	7
6.2	Challenges of making Sherpa standard language .....	8
6.3	Is the standardization of Sherpa language a dream? .....	9
7	Conclusion .....	9

### 1 Introduction

In Nepal there are more than 100 languages. Many of them are still under development, so these languages have not been yet analyzed in their phonology, orthography, and grammar. As a lexicographer the first job to make a dictionary for one of these language is the **dialect study**, to discover what dialects are in that language area. The decision on a more prestigious dialect will be helpful to take dialect variations out from the main entries. Furthermore, some languages, which are under two major different cultures like Sherpa: Indo Aryan and Tibetan civilization, need to consider one more topic of which **script** to use. Then the community must be involved in developing an **orthography** so that there will be a sense of ownership for literature produced.

For these decisions lexicographers should have field tests to see what the local people say. However, the problem is that in a society where tradition is so strong the conclusion of sociolinguistic study is not always the same as the whole society's opinion. The answers toward the linguistic survey may be scientific and practical, but for keeping

their identity the society may decide in a different way. So, the practical opinion and their philosophy to keep their tradition could not be the same. In this case lexicographer can face a real challenge. This paper is a report of the process of making a Sherpa dictionary.

## **2 The background of Sherpa language**

The Sherpa people (about 130,000 in the 2001 census of Nepal) live in the northern highlands of eastern Nepal and in nearby districts of China and India.

Originally the Sherpas were Tibetan, known as Kham people, but immigrated to Nepal across the Himalaya about five or six hundred years ago, because of political unstabilities at their homeland.

Sherpa language is classified within the Central Tibetan Subgroup of Tibetan Family of Sino-Tibetan (Voegelin and Voegelin 1965).

## **3 Dialect decision**

There are three main dialects in Sherpa area: East, West, and South dialect. In a sociolinguistic survey of the Sherpa area in 1998 (Lee 2003) some dialect distinctions were found, but strong differences were not noticed, which means these three dialects are very closely related. A Dialect Intelligibility Test showed that subjects from all three locations scored high on the Phaplu (Southern) text, and that the Southern dialect is recommended as the most understandable dialect for the speakers of the other two dialect (Table 1).

	SUBJECTS		
	West	South	North
West text	AV=95% N =11		
South text	AV=96% N =11	AV=97% N =11	AV=95% N =11
North text	AV=81% N =11		AV=96% N =10

**Table 1.** Sherpa dialect intelligibility (AV= average, N= number of subjects)

#### **4 Script survey**

In 2001 in preparing the Sherpa dictionary I conducted a survey on script preferences (whether Tibetan or Devanagari) for writing Sherpa language. I prepared two booklets of the same dictionary sample, in which the main entry was written in Tibetan and Devanagari scripts respectively. This test was done by face to face for 157 people, who were living in West, South, and North dialects speaking areas. The sample was deliberately skewed to over-represent the lamas (Buddhist religious teachers). Although lamas make up less than five percent of the Sherpa population, because lamas are very strong opinion leaders, and all of them are highly literate in reading and writing Tibetan, this survey included 16 lamas (ten percent) among 157 to hear more of their voices.

Even though this test was limited for dictionary purpose, 88% of whole subjects of the test preferred to use Devanagari. Only 3% of the non-lama subjects preferred Tibetan, as against 56% of the lama subjects. The result is as follows:

Subgroup	Dev. script	Tib. script	No preference	Total
Lama	7	9	0	16
Other	131	4	6	141
All	138 (88%)	13 (8%)	6 (4%)	157

**Table 2.** Preferences on script for Sherpa

On the basis of this result we decided in 2001 to use Devanagari script for the main lexeme entry, but include transliteration of the lexeme in IPA and representation in Tibetan script as well.

## **5 The practicality and Sherpa identity**

### **5.1 Challenges of Sherpa Dictionary project**

Nepali Constitution allows the minority language groups to teach their mother tongue until primary school level, and nowadays the government supports them to produce the primary school textbook in their minority languages. In December of 2003 the Sherpa Society, which is the top decision-making organization among the Sherpas, decided to produce Sherpa primary school textbook **in Tibetan script**, which means that Sherpa children in the future will be taught their own language in Tibetan script at school.

This script decision affects the Sherpa Dictionary project a lot. The main change is the script of the main entry from Devanagari to Tibetan. This script change is a great challenge to the lexicographer not only for an extra language learning (exactly speaking, Tibetan writing system), but also for the changes of sorting order of the main entry.

## **5.2 The practicality and Sherpa identity**

The percentage of Sherpa people literate in Nepali script is not precisely known, but the national literacy rate in 2001 was 65% for male, and 42.4% for female. And in Solukhumbu District, where the majority is Sherpa, the literacy rate was a little lower than the national rate: 56.6% for male, and 35.4% for female.

For the percentage of Sherpas literate in Tibetan script is also not known precisely, but could be similar to the numbers of lamas and ex-lamas among the whole Sherpas, maximum 5-10 %. Thus at least 90% of the whole Sherpa population cannot read Tibetan script. But the Sherpa community decided to teach Tibetan script for their children. What is the background of this decision?

It is well known that Tibetan identity rests on two pillars, Tibetan religion, and the Tibetan language written in Tibetan script. So, when people speak of Tibetan identity, it includes Tibetan script. Since most of the Sherpas are strong Tibetan Buddhists, they also believe that Tibetan script is holy, as their religion is holy.

After Tibetan script was developed in 7th century, the Tibetan writing system has remained unchanged for the last 1,300 years, even though oral Tibetan languages are quite different in each area. So, Tibetan people, if they are literate in Tibetan script, could communicate each other by their holy script, although their oral languages are so different. Therefore, Tibetan script is treated as holy to keep the whole Tibetan as one body. In this case if Sherpa community wants to be part of this great traditional heritage, it is essential for them to teach Tibetan script to their children. I think this is a major factor in the Sherpa identity.

### **5.3 Problem solving**

#### **5.3.1 Tibetan sorting**

If Tibetan script is used for the main entry, one of big problems is that the entries will be sorting by Tibetan order, and most Sherpas cannot read Tibetan, and therefore cannot find entries in the dictionary.

#### **5.3.2 Main entry in Devanagari with Tibetan transliteration**

Because of the problems of Tibetan sorting order, listing the main entry in Devanagari script would be an easier option for readers, with the Tibetan transliteration included after the IPA representation to give a guide to pronunciation. As most of the Sherpas literate in Tibetan are also literate in Devanagari script, this will be possible option. In this way the people who are only literate in Devanagari can use the dictionary without any problem.

#### **5.3.3 Index of Tibetan script in Appendix**

To provide for people who know only the Tibetan script, the dictionary will have an index of all entries written in Tibetan script together with Devanagari transliteration (which is the main entry of this dictionary), so that they can find the words through the index. In Appendix the index of all entries will come.

## **6 Standardization of the language**

### **6.1 Who does make the standardization of a language?**

The standardization of Korean language had not been completed until 1938. From 1906 Korea was under Japanese rule, and all official communication was forcefully changed into Japanese, so Korean language was really endangered. At that time some linguists had started to compile Korean dictionary to preserve Korean language and completed it in 1938. This dictionary was not only an opening door to develop Korean

language, but also the first standardization progress in dialect, alphabet and language as well.

The situation is not exactly the same, but Sherpa language is in danger to be extinguished with other Himalayan minority languages. There is no single book written in Sherpa language. No standard dialect and alphabet, neither standardization of the language yet. If the Korean experience could be an example in this situation, making a Sherpa dictionary is not only an opening door to develop the Sherpa language, but also the first standardization progress in Sherpa language. So, in minimum for the standardization, making a dictionary should be started immediately, simultaneously with other literacy and literature production project, but not waiting until too late.

## **6.2 Challenges of making Sherpa standard language**

Who does make the standard language of Sherpa? In choosing a standard alphabet there are two interest groups: the linguists, and the lamas. If the first is more on sound-based approach, the latter is based on the Tibetan writing system. For example, there are in Sherpa two sets of fricative: alveolar and alveo-palatal both with high and low register respectively. Linguists prepared two symbols with tone markers on each, whereas the lamas want four different symbols, because in Tibetan there are four different symbols.

Concerning choice of a standard **dialect** - as mentioned before, there are three dialects and the Southern dialect was decided as most widely understandable. However, there are strong arguments about this decision. The Northern dialect group insists that they are the first descendants who immigrated from Tibet and moved from the North to all the other places. Furthermore, the lamas who have religious leaderships are from the West and therefore promote the use of the Western dialect.

### **6.3 Is the standardization of Sherpa language a dream?**

Language standardization is not easy! The responsibility of the linguists is to seek scientifically the result (in orthography, etc) which is nearest to the people's real understanding and knowledge of **Sherpa**. On the other hand, the lamas have strong opinions based largely on their knowledge of how **Tibetan** works and is written, and they tend to reject results based on study only of Sherpa language use. Compromise with the lamas is essential. The Sherpa language does not belong to the linguists, but to the people who speak the language.

So, cooperation with the language society, in this case the Sherpa Society, becomes very important. The Sherpa Society, as a representative organization for all Sherpas, needs to take steps to formulate Sherpa language policy. The Sherpa Society should embrace both the practicality and Sherpa identity, and needs to decide on one standard Sherpa language for their future.

## **7 Conclusion**

More than five years' preparation for starting the Sherpa Dictionary is, I believe, worthwhile to wait for making a good dictionary. However, it is not wise to spend more time just for delaying the right decision for Sherpa language. In 1970 Burkhard Schoettelndreyer had produced 'A Devanagari spelling system for the Sherpa Language', but it was not used by anybody. For 34 years!! If the Sherpa Society does not make a timely decision on standardization issues, compromising where necessary, nobody knows when this opportunity comes again.

## References

- Central Bureau of Statistics. 2003. Statistical Year Book of Nepal. HMG National Planning Commission Secretariat
- Gordon, Kent. 1969. Sherpa phonemic summary. Uni. of Tribuvan & SIL.
- \_\_\_\_\_. 1970. Sherpa tone and higher levels. *Tibeto-Burman Linguistics* 3. 186-206
- Hale, Austin. 1970. A phonological survey of seven Bodic languages of Nepal. *Tibeto-Burman Linguistics* 3. 1-33
- Lee, Sang-Yong. 2003. A sociolinguistic survey of Sherpa. *Themes in Himalayan languages and linguistics*, ed. by Tej Ratna Kansakar and Mark Turin, 81-95.
- Schoettelndreyer, Heiderose. 1971. A guide to Sherpa tone. Uni. of Tribhuvan & SIL.
- Schoettelndreyer, Burkhard. 1970a. A Devanagari spelling system for the Sherpa Language. SIL.
- \_\_\_\_\_. 1970b. Vowels and tone in the Sherpa verb. SIL.