

The Construction of Sanskrit Database for a Polyglot Buddhist Dictionary

Jong-Cheol LEE

**Department of Philosophy and Religion,
The Academy of Korean Studies, Korea**

1. Outline of the SDICTP Project

"One of the most urgent tasks of Sanskrit lexicography lies without doubt in the field of Buddhism. When a hundred years ago Bohtlingk and Roth undertook to compile their great Sanskrit dictionary, editions of Buddhist texts were entirely lacking. (...) In 1899 the second edition of Monier-William's Sanskrit-English dictionary appeared. Additions of any importance to the Buddhist material contained in the small Petersburg dictionary do not occur in this work."

These words are found at the beginning of a paper entitled "Suggestions for a new Polyglot Buddhist Dictionary" which Professor J.W. de Jong published a half a century ago in the journal *Vak* (no.1, Poona:Deccan College, 1951). Professor de Jong suggested the compilation of a "Polyglot Buddhist Dictionary" as a step in the creation of a more complete Sanskrit dictionary. At the same time, he suggested ways to resolve several technical problems, such as the way in which to cite source texts. Unfortunately, the situation concerning Sanskrit dictionaries have not changed so much since Prof. de Jong made his suggestions 46 years ago.

Ever since 1992, when I was in India as a researcher, I have taken up Prof. de Jong's suggestions and am now operating the SDICTP Project (The Sanskrit Dictionary and Indian Classics Translation Project) with my own funds. This project was started in order to compile a Sanskrit-Korean Dictionary and a Polyglot Buddhist Dictionary. To lay the foundation for this project I have inputted (1) Indian Buddhist sutras and sastras and classics of grammar as well as major Sanskrit texts of Indian philosophy, and (2) Tibetan translations of several Yogacara Buddhist texts.

Originally the project was to have been composed of three parts: (1) input section, (2) section for compiling dictionaries, and (3) section for compiling critical editions. However, due to financial constraints, the plan was greatly reduced. Right now, I am the Director of the project, and there are three Indian Sanskrit scholars as co-workers. Sometimes I am also assisted by a Tibetan monk who inputs Tibetan translations.

2. Method for inputting Data

When inputting data, we must keep the following points in mind: (1) the compatibility of data, and (2) whether the data can be used with future operating system. With these two points in mind, we decided to use only the ASCII Code for inputting Sanskrit and Tibetan. All files are created using Text File. Sanskrit letters are inputted in the following way.

TRS	PCI
a	aa
i	ii
u	uu
r	R
r	RR
l	L
na	NGa
na	NYa
ta	Ta
tha	Tha
da	Da
dha	Dha
na	Na
sa	\$a
sa	Sa
ha	Ha
ma	Ma

*TRS: Transliteration in Roman letters, PCI: Personal Computer Input

*When transliterating Sanskrit letters into Roman letters, Visarga is h, and Anusvara is m.

*Avagraha is inputted using the transliteration form of "".

*Tibetan is inputted using the Sanskrit input form. For example, the Sanskrit 'nirvana' is 'mya nan las hdas pa' in Tibetan, but this term is inputted as 'mya NGan las Hdas pa'.

*Strong points of this form of inputting data include: (1) there is uniformity in inputting Sanskrit and Tibetan, and (2) compared to the KH(Kyoto-Harvard) system, it is easier to read the text in the form in which it is inputted.

3. Lists of Texts Already Inputted (See Appendix)

4. Future Plans

Because I attach great importance to the compatibility of data, I have mainly used the editor program and Tex (/Latex) when processing raw data.

When I decided to compile a Sanskrit dictionary, I had in mind a thick dictionary printed on sheets of paper. However, times have changed greatly. When we think of time wasted on proof-reading, revising and supplementing them, there is no need any more to think of dictionaries as printed works. It is time to seriously think of electronic dictionaries. If there is a way of creating databases directly from raw data, and if it is possible to search these databases freely, the original aim of a dictionary is fulfilled. With these demands in mind, I am considering a how to use the SGML--which possesses an environment similar to Tex, and which makes use of the tree structure and which allows databases to be transferred more easily than Tex-- as a midwife for the creation of an electronic dictionary.

In the future, I plan to create a electronic <critical edition of Abhidharmakosa> which shows the Sanskrit original text, Tibetan translation, Chinese translation and Korean translations in multiple columns. Now that fundamental databases have been completed, I think that those types of work is the first step to compile <polyglot buddhist dictionary>. But I would have to ask for assistance from a computer expert for this.

<Appendix>

catalog.txt

15-Jan-1998

Text name		bytes	Remark
<1>			
AKBH-VY1	SKT	226,593	sphutartha
AKBH-VY2	SKT	407,033	
AKBH-VY3	SKT	230,327	
AKBH-VY4	SKT	233,293	
AKBH-VY5	SKT	165,141	
AKBH-VY6	SKT	239,916	
AKBH-VY7	SKT	125,256	
AKBH-VY8	SKT	77,976	
AKBH-VY9	SKT	60,264	
AKBH	SKT	1,121,129	abhidharmakosa
<2>			
PV-V	SKT	1,067,130	prama navartikavrtti
HB-TIKA	SKT	412,282	hetubindutika
VN	SKT	72,762	vadanyaya
VN-TIKA	SKT	274,979	vadanyaya tika
NB-TIKA	SKT	174,741	nyayabindutika
<3>			
MANU	SKT	313,887	manusmrti
<4>			
SDP	SKT	540,713	saddharmapundarikasutra (Kem)
DBH	SKT	178,915	dasabhumikasutra (Rahder)
LANGKA-V	SKT	157,946	lankavatarasutra (Nanjo)

LANGKA-P	SKT	226,302	lankavatarasutra (Nanjo)
ASAA	SKT	1,641,736	Abhisamayalamkara (Wogihara)
BBH-2	SKT	446,676	bodhisattvabhumi (Wogihara)
BBH-1	SKT	60,404	bodhisattvabhumi (Rahder)
\$BH	SKT	459,906	sravakabhumi (Shukla)
<5>			
PP	SKT	576,102	prasannapada
<6>			
NBH	SKT	353,008	nyayabhasya
NS	SKT	33,088	nyayasutra
NVA-T	SKT	1,176,139	nyayavartikatatparyatika
NVA	SKT	673,610	nyayavartika
TBH	SKT	79,568	tarkabhasa
PRA\$ASTA	SKT	82,033	prasastapadabhasya
VS	SKT	21,226	vaisesikasutra
SPBH	SKT	348,138	samkhyapravacanabhasya
SS	SKT	27,473	samkhyasutra
STK	SKT	114,484	samkhyatattvakaumudi
YBH	SKT	130,362	yogabhasya
YS	SKT	12,323	yogasutra
<7>			
VY	TIB	716,883	vyakhyayukti (Tibet)
VYT-D	TIB	652,736	vyakhyayukti-tika (Tibet)
PSV-D	TIB	239,829	pratityasamutpadavyakhya (Tibet)
PSV-TIKA	TIB	706,327	pratityasamutpadavyakhya-tika (Tibet)
VKV	SKT	17,569	vimsatikavrtti
VK-D	TIB	3,301	vimsatika (Tibet)
VKV-D	TIB	22,231	vimsatikavrtti (Tibet)
TK	SKT	3,928	trimsika
TK-BH	SKT	68,105	trimsikavijnaptibhasya
TK-D	TIB	4,416	trimsika (Tibet)
KSP-D	TIB	40,425	karmasiddhi (Tibet)
MSABH	SKT	349,133	mahayanasutralamkara-bhaSya
DDVV	SKT	5,347	dharmadharmatavibhanga-vrtti

DDVV-D	TIB	41,790	dharmadharmatavibhanga-vrtti (Tibet)
MV-TIKA	SKT	334,445	madhyantavibhagatika
MVBH	SKT	68,523	madhyantavibhagabhasya
MVBH-D	TIB	98,314	madhyantavibhagabhasya (Tibet)
PSP-D	TIB	20,433	pancaskandha (Tib)
TSN	SKT	4,495	trisvabhavanirdesa
TSN-D	TIB	5,544	trisvabhavanirdesa (Tib)

<8>

MBHA	TXT	2,449,606	Mahabhasya
SIDHK	SKT	1,089,903	Vaiyakaranasiddhantakaumudi
KAŚIKA	TXT	1,773,871	Kasikavrtti

<9>

TATTVAS	SKT	1,361,817	Tattvasamgraha (+panjika)
---------	-----	-----------	---------------------------