

## **Comparative Dictionary of Tibetan Dialects (CDTD)** **A research report**

by Roland Bielmeier

This dictionary is the outcome of more than 10 years of project work. Most of the data were collected by project members in the field, other data were taken from recent and reliable publications. We started our first project, financed by the Swiss National Foundation, in 1992 with the title ‘Historical-Comparative Lexicon of Tibetan Dialects’. It ended in 1995 and the Swiss National Foundation agreed to support a second project, ‘Foundations of a Historical Grammar of Tibetan’, which lasted with some breaks till 2000.

### **Goals**

A comparison of the lexicon of Written Tibetan, i.e. of the classical and pre-classical language, with that of its modern spoken dialects shows that the latter have undergone a simplification of the syllable structure. However, the nature and degree of this simplification varies between the different areas in which Tibetan is spoken as a mother tongue. The documentation of this historical language change in Tibetan on the phonetic-phonemic and lexical levels constituted the goal of our first project. To work towards this goal, we set out to collect a limited selection of vocabulary items in as many spoken varieties as possible, and to compare this vocabulary with the etymologically corresponding lexicon of Written Tibetan. Our main aim was to discover the principles of sound change in Tibetan over the last twelve centuries. The second project was linked to the preceding project in terms of both data and goal. The goal was, on the one hand, to continue our work with the dictionary and, on the other hand, to extend our research to all linguistic levels. In the first project we had dealt mainly with the phonetic-phonemic and lexical levels from a diachronic point of view. Now grammar was to be included in the diachronic approach.

The basic methodological idea remained the same, i.e. to facilitate a comparison between the grammars of spoken and written Tibetan. We selected a number of spoken varieties for which to write grammars, basing our choice on two criteria. First, we intended eventually to produce reliable grammars from all main areas of linguistic Tibet. Second, archaic or conservative varieties were to be given first priority. Our further aims are to establish a new genetic classification of the Tibetan dialects, based mainly on the principles of sound change, and to establish the dialect geography of linguistic Tibet, based mainly on sound and lexical change.

### **Research group and fieldwork**

In Western Tibet the fieldwork was conducted by myself and Ngawang Tsering. I had already conducted fieldwork on Balti and Ladakhi between 1986 and 1988 and continued this work after the beginning of the first project in 1992. In this I was supported by Ngawang Tsering, a native speaker of the Lower Ladakhi dialect of Nurla. Ngawang Tsering had collected an impressive amount of dialectal data on the different varieties all over Ladakh, especially between 1995 and 1997, and we transcribed and evaluated this data together. Felix Haller and Chungda Haller, a native speaker of Shigatse Tibetan, were also members of the research group for the first project and they conducted fieldwork in Central and Eastern Tibet. They visited Shigatse three times to collect material, but our main informant for the Shigatse dialect is Chungda Haller. Felix Haller finished his Ph.D. thesis on the Shigatse dialect in 1995. This comprehensive grammatical description, including transcribed and translated narratives, was published in 2000.

In addition, they conducted fieldwork in Eastern Tibet, Amdo and Kham, from 1992 to 1997. The first objective of this fieldwork was to obtain a general overview of the dialects spoken in Eastern Tibet. In 1992 they were in Amdo collecting material from Themchen, Kangtsha, Kharmar, Chapcha, Dzorganrawar, Matö, Labrang and Ngawa. The second fieldtrip in 1993 took them to Kham, where they collected material from Dartsedo, Kardze, Derge, Lithang and Bathang. After these two fieldtrips Felix Haller concentrated on the highly conservative North Amdo dialect of Themchen which enjoys great prestige in Amdo. Mostly during the summers from 1994 to 1997 Felix and Chungda Haller collected material on this dialect. Because Themchen was a closed area at that time, they visited it only twice and decided to do most of the fieldwork in Xining, asking their informants to come there. In 1995 Katrin Häsler, Veronika Hein, Brigitte Huber and Marianne Volkart joined the second project. In cooperation with the South-West College of Nationalities in Chengdu, Katrin Häsler conducted two fieldtrips to Kham in 1995 and 1996 and in 1999 presented a grammar of the Tibetan Dege dialect, which also enjoys high prestige in Kham, as a Ph.D. thesis. Part of the recorded material was transcribed and analysed during the fieldtrips. The majority of the transcription and analysis was, however, conducted in Switzerland with the help of language informants from the Dege area. Veronika Hein is working on the Spiti variety of Tibetan in Tabo in Northern India. Since 1995 she has made a fieldtrip to this area every year and will complete her Ph.D. thesis on the Spiti dialect soon. Brigitte Huber conducted almost a year of fieldwork in Nepal (in 1998, 1999, 2000 and 2001) working on the dialect of Kyirong, north of Kathmandu. As she did not receive permission to travel to Kyirong on the Tibetan side of the border, she worked with Kyirong speakers in Nepal. In 2002 she finished her Ph.D. thesis on the Lende subdialect: this consists of a description of the dialect and also presents a historical perspective by comparing it with Written Tibetan.

Some of the material included in the CDTD was taken from published sources on the Central Tibetan varieties of the Western Drokpas and of Southern Mustang (Monika Kretschmar 1986, 1995), of Dingri (Silke Herrmann 1989), from sources on the Kham Tibetan variety of Nangchen (Margret Causemann 1989), and on the Amdo Tibetan variety of Ndzorge (Jackson T.-S. Sun 1986). These data were entered into the computer during the first project. During the second project Marianne Volkart entered material of the Jirel in Nepal (Anita Maibaum and Esther Strahm 1971, 1973, 1975), while Chungda and Felix Haller entered the Bathang material of the Tibeto-Burman lexicon (Dai Qingxia et al. 1992) provided by Kesang Gyurme who is a native speaker of Bathang Tibetan; the Arik material of the Tibeto-Burman Lexicon; the material of seven Ngari Tibetan varieties complemented by Lhasa Tibetan (Qu Aitang and Tan Kerang 1983); the Dzongkha material of George van Driem's grammar of Dzongkha (1998); and the Golok material from Serta (Richard Keith Sprigg 1968, 1972, 1979).

### **Processing of the data**

The fieldwork data were recorded on tape, transcribed by developing the phonemic analysis of the different varieties, and entered into the computer. With regard to the verbs, simple sentences were elicited and entered into our database together with notes on the verb classification documented by the sentences. In the case of published sources, usually all the available material was taken and phonemic transcriptions were adapted to our system. An English or German translation was added to all entries, as well as a reference to the data source, and the German translation was complemented by an English correspondence.

The last step in processing the data was aimed at establishing the internal Tibetan etymology. This was achieved by applying the historical-comparative method, well known from Indo-European studies, to establish regular sound correspondences between the orthographic forms of Written Tibetan and the spoken dialectal forms. As the data for the dictionary grew, we had to reorganise our database completely, an undertaking which was headed by Katrin Häsler with the help of Moritz Vögeli, who designed the new 'Toblerone' fonts for the dialectal and Written words. She succeeded in simplifying the structure of the database. The most important change was the result of her idea to assign a specific key number to every Written Tibetan lemma, a number that is also assigned to all the etymological correspondences in the different varieties, which are finally to be grouped together under the Written Tibetan lemma. Thus we were able to escape the problem that the slightest difference in word formation would lead to chaos in our database. On the other hand, this new strategy brought us back immediately to the still unresolved problem of how to handle Written Tibetan. We finally decided to build up two data stacks, one containing the verbs of Written Tibetan, aiming at completeness, and one containing all those Written Tibetan nouns, adjectives, numbers, etc. for which we

had dialectal evidence. The Written Tibetan forms are based on H. A. Jäschke's Tibetan-English dictionary of 1881 and its German precursor of 1871, on Goldstein's Tibetan-English Dictionary of 1978 [2001], and on the Great Tibetan-Chinese Dictionary in three volumes of 1985 (Zhang Yisun et al., *Bod Rgya tshig mdzod chen mo. Zang Han da cidian*. [The Great Tibetan-Chinese Dictionary]. Vols. 1-3. Beijing 1985). Both the verb stack and the noun stack of Written Tibetan were entrusted to Marianne Volkart, who is writing a Ph.D. thesis on word formation in Written Tibetan. The highly labour-intensive task of assigning the correct etymology to every dialect item was shared among the members of the research team. Basically the fieldworkers were responsible for 'their' dialects, whereas the data from the published sources were dealt with mainly by Marianne Volkart, Brigitte Huber and myself. A further important organisational change, concerning the editing of our dictionary, was the decision to plan the final export of the complete data not into Word, which is unable to process such a quantity of data, but into QuarkXPress.

### **The dictionary**

It has been our aim from the beginning to finally make our data accessible in the form of a dictionary modelled on R. L. Turner's *A Comparative Dictionary of the Indo-Aryan Languages*, (Oxford 1966/1969). Following Turner, who arranged all the Middle and New Indo-Aryan words under the etymologically corresponding Sanskrit word, we decided to arrange all the Tibetan dialectal words according to a new dialect classification under the etymologically corresponding Written Tibetan words, arranged alphabetically according to the traditional Tibetan system. And it was in these early stages of the second project that we decided (still following Turner to a certain extent) to call our dictionary the *Comparative Dictionary of Tibetan Dialects* (CDTD).

Two preprints of the verb volume of the CDTD were printed in September 1997 (main volume 253 pp. in two columns, with 1532 main entries of Written Tibetan verbs and 32 dialectal varieties) and September 1998 (main volume 272 pp. in two columns, with 1567 main entries of Written Tibetan verbs and 34 dialectal varieties). These preprints include an introductory section with a preliminary new classification of Tibetan dialects, the phonemic inventories, etc., and to present our results to the international community on the occasion of the 8th Himalayan Languages Symposium we printed a third preprint of the verb volume in September 2002. The main volume (325 pp. in two columns) contains 1602 main entries of Written Tibetan verbs with their dialectal correspondences of 67 varieties. This volume is accompanied by two index volumes. The first contains an index of all Written Tibetan words, as well as the indexes of the entries for each dialect (209 pp.). The second contains the indexes of the English and German translations and gives for each word the lexical type based on the Written Tibetan entry, as well as a list of the dialectal varieties in which the word is documented (199 pp.) The first two preprints

of the noun volume of the CDTD including the index volumes are organised in the same manner as the verb volume. They were printed in April 2000 (main volume 557 pp. in two columns, 5710 main entries of Written Tibetan nouns, 57 dialectal varieties) and July 2001 (main volume 646 pp. in two columns, 6006 main entries of Written Tibetan nouns, 67 dialectal varieties). As we could not finish the current revision of the noun volume till the 8th Himalayan Languages Symposium, we printed as third preprint only the first revised half of the main noun volume together with the corresponding index volumes. Due to the growing number of pages especially of the noun volume we now printed some of the indexes separately. This may be changed in the final version.

Currently, we are working on the following volumes:

Volume 1 : Introduction (research history, phoneme inventories, principles of presentation, classification of the verbs, classification of the Tibetan dialects and its criteria, geographic features of Tibetan dialects, etc.)

Volume 2.1 : Noun volume

Volume 2.2.: Written Tibetan index

Volume 2.3 : Dialect indexes

Volume 2.4 : English index

Volume 2.5 : German index

Volume 3.1 : Verb volume

Volume 3.2 : Written Tibetan and Dialect indexes

Volume 3.3 : English and German indexes

For further details see <http://www.isw.unibe.ch/tibet/>

Roland Bielmeier

University of Berne (Switzerland)

roland.bielmeier@isw.unibe.ch