AUSTROASIATIC AND TAI-KADAI LANGUAGES IN THE INTERCONTINENTAL DICTIONARY SERIES

Commrie Bernard, Periros Ilia

The study of the lexical diversity of human languages is, probably, one of the most exciting aspects of linguistics. However, it is also one of the most time consuming and complex parts of the discipline. Even very simple tasks require extraordinary efforts and could take a long time to complete.

Let us consider a typical example. In the process of researching into the cultural development of Mainland Southeast Asia, one would need to determine what is the word for a particular animal, for example - pig, in approximately two hundred local languages. To obtain this information, one could use several etymological or comparative dictionaries of these languages. These dictionaries, however, would only reveal details of words with reliable etymologies, while all other words would, naturally, be absent from this particular type of source. The rest of the words required for the research would, then, be obtained from various bilingual dictionaries, where forms are usually alphabetized according to the language in question, while the meanings are given in English, Chinese, Vietnamese, Thai, French, Dutch, Russian, German or Indonesian. Most of these dictionaries would use their own phonological representations with different notations for tones, vowel length and other features. Therefore, it would take approximately two weeks of intensive (and not very productive) work to get the information related to the meaning 'pig'.

The situation is quite different for the Tibeto-Burman languages of China. With the aid of "A Tibeto-Burman Lexicon" edited by Huang Bufan (1992), one could get the same information for fifty-one languages and dialects, in only a few minutes. This publication brought together words from different languages with identical meanings into a simply and concisely structured arrangement. Therefore, to establish the representations of the meaning 'pig' in some Tibeto-Burman languages, all one needs to do is simply open the dictionary and read the information given.

With no generally accepted term, one could call such a dictionary 'comparative synonym dictionary'. It is comparative, as it provides data from various languages and it is also a dictionary of synonyms, as it lists forms from different languages with identical meanings.

Huang Bufan's book is not the only published comparative synonym dictionary (for the Tibeto-Burman languages see, for example Hale 1973 or Sun 1991) and is, in fact, a continuation of a long linguistic tradition. However, printed comparative-synonym dictionaries are complex to use, and their rich substance is often not fully exploited. Therefore, over the past years, general interest in such dictionaries has not been exceptionally strong.

Modern electronic databases have given a new direction to comparative synonym dictionaries and this opportunity is used in the Intercontinental Dictionary Series (IDS) project. Marie Ritchie Kay (University of California, Irvine) has developed this project with the aim of establishing a database, which would contain lexical material from
languages of the world organized in such a way as to allow comparisons to be made simply.

An entry on the IDS database will reveal how a particular meaning is conveyed in the selected languages. Therefore, an entry is, in fact, a list of synonyms found in the different languages, thus making the database a typical comparative synonym dictionary. The final outcome of the project will be an electronic database representing the entire range of human languages. It will bring together information on languages of the world published in dozens of different languages and scripts and scattered in hundreds of publications and manuscripts, which are often not available to the linguistic community. This database will also include word-lists of less known languages collected especially for this project.

Currently the IDS project is predicted to be a set of interlinked electronic databases representing various language families or geographic regions. We are hopeful that eventually the IDS databases will contain most, if not all the languages in the world. Clearly, then, the IDS project is an enormous task, which may be achieved only through wide international cooperation among linguistic institutes and individual scholars.

Apart from being an archive of lexical data, the IDS database will also provide rich resources for various studies such as:

(i) evaluation of the linguistic diversity of regions and thus, for developing detailed and justifiable linguistic maps;

(ii) genetic and non-genetic classifications of languages;

(iii) typological studies based on lexicon, such as analysis of sound symbolism, word formation, phonological theory;

(iv) etymological and contact studies, such as identification of contact zones, an analysis of the spread of various cultural ideas, or proto-lexicons.

The IDS project is based on a list of meanings, which largely follows that used by Buck (1949) in his *A Dictionary of Selected Synonyms in the Principal Indo-European Languages*. Buck's dictionary contains approximately 1200 meanings, which, together with their numbering systems are incorporated into the IDS master list. Approximately 100 additional meanings have been added to it by M. R. Key.

Therefore this master list (1,310 entries) consists of:

1. universal meanings found in most human languages (HAND, SPEAK, DRY, etc.)

2. meanings related to certain geographical or environmental phenomena: EARTHQUAKE, TIDE, PARROT, etc.

3. meanings reflecting certain cultural ideas: MEAD, TATTOO, COBBLER, etc.

Naturally, not all meanings from groups (ii) and (iii) will be found in all the languages.

In some cases, it may be important to add extra meanings to represent information relevant to a particular language group or a region. Therefore, in the process of data collection, the list of the meanings will be expanded, but under no circumstances will the meanings be lost from the master list. If the corresponding meaning is not known in a particular language, the entry will simply remain blank. The reason for this strict rule is quite clear: by adding meanings, the main body of the databases is still kept compatible, while deleting meanings may create databases which are no longer compatible.
Several dictionaries based on the general principles of IDS have already been prepared, or are in preparation, including the *Comparative Austronesian Dictionary* edited by D. Tryon et al. (1995) and the *South American Dictionary* edited by M. R. Key.

One of the IDS projects which is currently in preparation is the "Austroasiatic and Tai-Kadai comparative synonym dictionary". Its aim is to represent in the IDS format all the languages of these two families. At this stage, however, we are not including the Munda and Nicobarese branches of Austroasiatic.

The languages of these two groups are often spoken in the same areas of China, Thailand, Vietnam, Laos, Cambodia, Malaysia, Burma and India. As these languages also often reveal traces of intensive and sometimes old mutual contacts it is preferable to study the two families in the same project.

The preliminary work on the project includes a creation of the lists of the languages to be investigated and some additions to the IDS list of meanings. This modified list of meanings is to be found in the appendix.

A much more challenging task has been to compile a list of languages or dialects to be included in the investigation. With the exception of Myanmar (Burma), territories where Austroasiatic and Tai-Kadai languages are spoken are now open to linguistic research. This means, that new languages are constantly being brought to light, and that there is a significant chance that new discoveries will be made in all of these countries. On the other hand, these two linguistic families are not very popular among linguists and in some cases, a language has not been included in the project simply for the reason that no linguist was found to conduct the research.

The list of languages already included in the project is given below. This list has been compiled with the inestimable help of colleagues: G. Diffloth, M. Ferlus, Li Jinfang, Luo Yongxian, S. Morey, Nguyen Van Loi, Sun Hongkai, Suwilai Premsrirat, Theraphan L. Thongkum, Wilaiwan Khanittanan and Zhang Qiusheng.

Most of the Tai-Kadai languages are known relatively well, with the basic information lacking only for some members of the Ge-Yang branch. Therefore, the main goal of the Tai-Kadai part of this project is to focus on the finer details of the bigger picture.

The central role in the Tai-Kadai family is played by the Zhuang-Tai languages, which are divided into Shan-Tai (South-Western), Nung (Central) and Zhuang (Northern) branches.

The largest area associated with the Shan-Tai languages covers Thailand and Laos. Both of these countries have their own official languages (Thai and Lao), which will be included in the project. The linguistic situation here, however, is much more complex, as some dialects of Thailand have differences comparable to those observed between Thai and Lao. Therefore, the project will include six dialects from Thailand, which, hopefully, will represent the linguistic diversity of this country:

1. Central = Thai
2. Khamuang (Chiang Mai)
3. Southern Tai
4. Dialect of Nongkhai
5. Dialect of Khorat
6. Eastern Thai (Trat)

At this stage the Lao language is represented only by
7. The dialect of Luang Prabang.
The dialect of Vientiane is probably very close to that of Nongkhai.
The Shan-Tai languages of India are represented in the project by:
8. Khamti
9. Aiton
Aiton is a representative of four quite similar and mutually intelligible dialects spoken in Assam.
The Shan-Tai of Burma are represented by:
   - two varieties of Shan:
     10. Northern Shan and
     11. Southern Shan, which is known in Thailand as Tai Yai
The Shan-Tai languages of China are spoken only in Yunnan. For this region, it has been decided that the data should be collected in the same way as in Thailand: covering the main locations where these languages / dialects are spoken:
13. Dehong (Chinese Dai)
14. Lue (Sipsonpanna Dai)
15. Jinsha Dai (probably the most northern dialect of Shan-Tao)
16. Jingping Dai
17. Tai Ya
Currently it is rather difficult to evaluate the linguistic diversity of these languages / dialects.
The list of the Shan-Tai languages of Vietnam to be included in the project has not been finalized yet. So far, it consists of:
18. White Tai
19. Black Tai
20. Red Tai
21. Tai Muong
22. Tai Nam
The Nung or Central dialects of Zhuang-Tai are spoken mainly in Vietnam and China, where they are now called “Southern Zhuang” languages. Five varieties (three from Guangxi, China and two from Vietnam) have been selected for the project:
23. Longzhou Nung, which is, probably, the best know representative of this group;
24. Debao Nung
25. Lazhai Nung
26. Nung Fan Slihng
27. Tay Nung
We are hopeful that these five dialects will represent the entire diversity of the Nung group.
Three dialects (languages) have been chosen for the Zhuang group:
28. Saek of Thailand / Laos border