THE INDEXING OF CHINESE NAMES

H. D. TALBOT

All indexers find themselves out of their depth at some time or another. Probably the indexing of oriental names is a field where most indexers will feel uncertain of themselves, and would not object to some guidance. The following account is meant to indicate the main problems encountered in the indexing of Chinese names, and where the information may be found to solve them.

The Indexing of Chinese Characters

Chinese is not written in an alphabetical script, but by a series of lines (strokes) which together form a character. This character will convey an idea, more or less definite, have a pronunciation, which is usually constant in a given dialect, and may be considered as a word by itself or part of a word when combined with one or more other characters. There are something like fifty thousand of these characters, although not all these are in common use.

The need for dictionaries was realized by the Chinese at a very early date, and methods were invented to index the characters. One method was to decide upon 540 combinations of strokes, called radicals in English, one of which could be recognized in each character, and to arrange the characters in groups according to which radical they contain. The radicals themselves have a definite order, as they are arranged in groups according to the number of strokes they contain. All radicals of one stroke come together first, then those of two strokes and so on. In the 18th century the number of radicals was reduced to 214, and this is the number used in modern Chinese to English Dictionaries.

Another traditional method of indexing Chinese was the “Rhyme Dictionary” arranged on phonetic principles, but as the pronunciation of Chinese has changed since these were first made, and the arrangement of the dictionaries did not change at the same time, this method became somewhat difficult to use.

In modern times new characters have been invented—for instance, to stand for the chemical elements; and at the same time many characters have been abbreviated. It was realized that the radical system was too cumbersome and a new system was needed.

The “four-corner system”, which assigns a number from 0 to 9 for each corner, according to the way the corner is written, answered this need. In this system a character will be described by a four-figure number with one extra figure added in small type. The characters are then arranged in numerical order.

A particularly interesting system of indexing Chinese is that used by the Russians. It simply uses the morphology of the characters: all characters with predominantly horizontal strokes are grouped together and so on. It might well be a convenient system to use with the scanning devices of an automatic translation system.
The Transcription of Chinese

Ever since the first Portuguese contracts with China in the 16th century, attempts have been made to transcribe the sound of spoken Chinese into the Latin alphabet. As would be expected, the first attempts were unsystematic, and gradually systems arose which depended on the mother language of the Europeans concerned.

There were great difficulties to be overcome in inventing these systems, which were caused by the nature of the Chinese colloquial language itself.

Firstly, not all the Chinese speak their language with the same pronunciation, although the written language is the same in all parts of China. In the south, and along the south-east coast up to the mouth of the Yangtse, there are a number of distinct dialects, which the Europeans first encountered and which are widespread, as most emigrant Chinese speak one of them. This indeed is the explanation for the strange names on menus in Chinese restaurants in London when compared with the more usual transcriptions of the "Mandarin language" based on the Peking dialect, which is most widespread in China itself.

Even when the Mandarin is considered for purposes of transcription, the matter is by no means simple. Although there are only 412 distinct syllables in the Mandarin, many of the sounds are not found in any European language, and each one can be said with one of four tones, which give it a completely different meaning. An idea of the importance of tones can be gained by remembering the different ways of saying the word "yes" in English, to mean interrogation, enthusiastic assent, incredulity and so on. In Mandarin "Fu" can mean "a distinguished person", "to help", "to bow down" and "near to", according to the tone.

The system of transcription which has become standard for British official purposes was initiated by T. F. Wade in 1859 and modified by H. A. Giles in the two editions of his dictionary. This system is therefore known as the Wade-Giles system.

From the point of view of indexing it is important to note that the initial consonant or consonants in the Wade-Giles system may be followed by an apostrophe, which modifies the pronunciation. Thus "Tang" and "T'ang", "Tsa" and "Ts'a", must be kept distinct in indexing.

The tones are indicated by numbers in small type after the syllables, and thus a further distinction can be made in indexing.

Chinese Personal Names

Each person has a surname (Hsing) and a personal name (Ming). Usually the whole name (Hsing+Ming) is written in three characters, the surname being a single character (therefore a single syllable), and the personal name is in two characters. However, double-syllable surnames are possible, and many-syll-
able surnames of people of Manchu and Mongolian origin also occur. Normally the surname is given first, and if it is of the single-syllable type it will not be hyphenated to the next syllable. If a double-syllable surname is suspected, the only thing to do is check it against the list of surnames on pages one to eight of Giles' dictionary, and in the reference books if the person is likely to be included. The only certain way of identifying the surname is to have the whole name in characters.

The two syllables of the personal name are conventionally written in transcription hyphenated together. This is the legal name and the one signed on legal occasions.

However, a person may be known by many other names as well during the various phases of his life, and this renders the identification of people difficult when these names are used in Chinese literature.

To sum up, the usual form is:

Ku Tung-kao

where "Ku" is the surname.

A form which sometimes occurs is:

Hu wei

where "Hu" is the surname.

Another possible form is:

Ssu-ma Ch’ien

where "Ssu-ma" is the surname.

**Chinese Imperial Titles**

Each emperor had a surname and personal names, but during the dynasty they were forbidden to be uttered or written by ordinary mortals. Therefore an emperor was designated in other ways. He would have a posthumous title (Shih) which might be very long, but was cut down to one or two words. Also a posthumous "Temple Name" (Miao-hao); these tended to recur in different dynasties, and so they are written preceded by the name of the dynasty.

The usual name used in European literature is strictly a "Reign Title" (Nien-hao), which was only used for whole reigns after 1368. Names such as K’ang-hsi and Chien-lung are of this type, and should be the main entry in indexing.

**Chinese Titles and Ranks**

These are extremely unlikely to occur in such a way as to be mistaken for part of the name; they are usually translated in European books.

**Chinese Place Names**

The problem of transcription occurs here as well. The standard British list of Chinese names of the nineteenth century (Playfair) used the Wade system, but
this has become superseded by the “Postal Guide” system, which was adapted to local dialects, and even contained intentional mis-spellings to distinguish places with similar names.

The main problem encountered in Chinese place names in the text of a book usually stems from an author's lack of understanding of the names. It is very often not realized that the name itself contains the word “mountain” or “river” in it, so there is no need to repeat it in English. For instance, in a recent geography textbook the author used the names “Yangtse”, “Yangtse River”, “Yangtse Kiang” and “Yangtse Kiang River”. The last form is erroneous as “Kiang” means river. The name was indexed as “Yangtse Kiang”, the full Chinese form. Similarly, “Shan” means “mountains”, and to say “Nan Mountains” instead of “Nan Shan” is somewhat absurd, when it is realized that “Nan” means “south”. The rule must be translate all or nothing.

Another thing about Chinese place names which might well mislead an indexer are the suffixes on Chinese town names. Up to about 1911 each Chinese town had a definite administrative rank, and this was indicated by the suffixes which were (in descending order) Fu, T'ing, Chow or Chou, and Hsien or hien. These suffixes are sometimes missed out, and then the indexer may believe he is dealing with different places. For instance an author may refer to “Sining” and “Siningfu” indiscriminately.

Unfortunately, these suffix syllables can be an integral part of the name, so that “Soochow” is not complete and the full form is “Soochowfu”. The only safeguard here is careful checking in the reference books.

**Further Sources of Information**

**The Indexing of Chinese Characters**


**The Transcription of Chinese**


**Chinese Personal Names**


CORRESPONDENCE

INDEXING A CLASSIFICATION SCHEME

Dear Mr. Mills,

In the documentation profession as elsewhere there are fashions. One of the present fashions is chain indexing, which you so clearly describe in your article in the autumn number of *The Indexer*. Reading this from the point of view of the layman I am moved to ask a few questions and to make a few comments.

On the assumption that this form of indexing were attached to a Comprehensive bibliography your suggestion on page 43 that the absence of an entry on Painting, Landscape 758.1 will not prevent the enquirer finding material at “Painting 750” seems to me to hold a fallacy. If there are two or three pages of entries under painting generally he will never move to 758.1. Why should he, since 751 “material and methods” will lead him to think he has passed to a new field? This is no doubt a weakness in Dewey but an index must allow for this.