



## United States Board on Geographic Names

### Word Segmentation Standard for Chinese Geographic Names

Approved by 325 FNC, July 10 2007

#### Scope

The following policy applies to all geographic names in the Chinese language, including the names of geographic features in the People's Republic of China, including Taiwan, Macau, Hong Kong, as well as Chinese-language names from other regions of the world. The policy describes how to segment strings of Chinese characters into separate words during romanization. For the purposes of this document, only simplified Chinese characters are shown, even though traditional Chinese characters are used in certain areas. This policy shall not be taken to imply that the modified Wade-Giles system of romanization is no longer approved for Taiwan by the BGN.

#### Related Documents

The document *Romanization Systems and Roman Script Spelling Conventions* (Defense Mapping Agency, 1994, republished 2006) describes general rules for Chinese Romanization.

For each of the countries and territories mentioned above, a Geographic Names Standardization Policy (GNSP), available from the United States Board on

Geographic Names (BGN), describes country-specific aspects of romanization, capitalization, hyphenation, and other technical treatment.

For each of the countries and territories mentioned above, a country-specific generic terms glossary lists generic terms recognized by the BGN.

## Policy

1. For all Chinese-language geographical names, regardless of country of origin, the BGN approved forms follow the rules for word segmentation found in the document “Basic Rules for Hanyu Pinyin Orthography,” sections 4.1 and 4.2. This document represents the national word-segmentation standard of the People’s Republic of China (ICS 01.140.10), numbered GB/T 16159-1996, effective July 1, 1996.

2. Summary of Rules

- a. When a geographic name consists of a generic element and a specific element, the generic and specific elements shall be separated by a space in romanization, except as amended by rule 2.b and note 1 below; for example:

*Xiping Mountain*

西平山

Xiping Shan (HY)<sup>1</sup>

Hsi-p'ing Shan (WG)

The first letter of the generic and the specific element shall be capitalized in romanization.

- b. For names of populated places (PPL, PPLX, PPLA, etc.) or less, the generic and specific elements are written together as one word. This rule includes such terms as *cun*, *keng*, *zhen*, etc.; for example:

泉州坑

Quanzhoukeng (HY)

Ch'uan-chou-k'eng (WG)

---

<sup>1</sup> HY = Hanyu pinyin romanization; WG = Wade-Giles romanization.

- c. Monosyllabic<sup>2</sup> adjuncts<sup>3</sup> to the generic term are written together as one word with the generic element; for example:

*Baigou South Mountain*

白狗南山

Baigou Nanshan (HY)

Pai-kou Nan-shan (WG)

- d. Monosyllabic adjuncts to the specific element are written together as one word with the specific element; for example:

*South Street inside Chaoyangmen Gate*

朝阳门内南小街

Chaoyangmennei Nanxiaojie (HY)

Ch'ao-yang-men-nei Nan-hsiao-chieh (WG)

- e. If the generic or specific element of a geographic name consists of more than one integral concept (cí 词), it shall be further segmented on the basis of word boundaries or juncture. If this is not possible, then the term should be written as one word.

- i. Example structures of two or three syllables that express an integral concept that are to be written together as one word:

*College, University*

大学

Daxue (HY), Ta-hsüeh (WG)

*Museum*

博物馆

Bowuguan (HY), Po-wu-kuan (WG),

*Farm*

农场

Nongchang (HY), Nung-ch'ang (WG),

---

<sup>2</sup> i.e. one syllable.

<sup>3</sup> i.e. modifiers.

- ii. Example structures of four or more syllables where word segmentation is necessary:

*Agricultural Improvement Farm*

农业改良场

Nongye Gailiangchang (HY)

Nung-ye Kai-liang-ch'ang (WG)

*Chinese Academy of Social Sciences*

中国社会科学院

Zhongguo Shehui Kexueyuan (HY)

Chung-kuo She-hui K'o-hsüeh-yüan (WG)

- iii. Example where word segmentation is not possible without changing meaning or ungrammaticality:

*Graduate School*

研究生院

Yanjiushengyuan(HY)

Yen-chiu-sheng-yüan (WG)

For the purposes of this policy, if a sequence of characters is listed as a separate entry in the dictionary *Xīnhuā Zìdiǎn* (2004) or in the current country-specific geographic glossary, and is not a one-syllable adjunct as in rules 2.c and 2.d, it is considered an integral concept. This statement shall not be construed as negating any of the other rules contained in this policy.

- f. If the specific element of a geographical name itself contains the name of another geographical feature, the specific name is segmented in the same way that the embedded feature's name would be. For example:

*Chaoyang District Agricultural Improvement Farm*

朝阳区农业改良场

Chaoyang Qu Nongye Gailiangchang (HY)

Ch'ao-yang Ch'u Nung-ye Kai-liang-ch'ang (WG)

*Qingdao Harbor Industrial Area*

青岛港关连工业区      Qingdao Gang Guanlian Gongyequ (HY)  
Ch'ing-tao Kang Kuan-lien Kung-yeh-ch'ü (WG)

*Shuilian South Stream Bridge*

水琏南溪桥              Shuilian Nanxi Qiao (HY)  
Shui-lien Nan-hsi Ch'iao (WG)

**Notes:**

1. Occasionally the generic element is embedded between two parts of a split specific element. In this case, a space shall be inserted in romanization on either side of the generic element; for example:

*Guangdong Provincial Xinying **Hospital**, Beimen Branch*

广东省立新营医院北门分院

*Guangdong Shengli Xinying **Yiyuan** Beimen Fenyuan (HY)*

*(Generic in **bold**, specific in italic)*