Etymology in Taiwanese

Henning Klöter
Leiden University
Research School CNWS
P.O. Box 9515
NL-2300 RA Leiden
The Netherlands
H.Kloeter@let.leidenuniv.nl

Abstract

Since the early 1930s, Taiwanese scholars have conducted research on the etymologies of native Taiwanese expressions. Their philological investigations have resulted in a number of dictionaries. A comparison of these dictionaries reveals a high degree of orthographic divergence. Section 1 of this article provides a short overview on the linguistic make-up of Taiwan, followed by an introduction on Taiwanese lexicography and orthography. In Section 2, the earliest proposals for an etymological Taiwanese orthography are introduced, and compared with solutions in dictionaries published since the late 1990s. The last section discusses the import of etymological considerations for the orthographic standardization of Taiwanese.

1 Introduction

1.1 The linguistic make-up of Taiwan

Taiwan's turbulent history has brought about a complex sociolinguistic make-up. Located off the southeast coast of China, the island of Taiwan was once exclusively inhabited by speakers of Austronesian languages. Chinese immigrants, mainly from the opposite province of Fujian, began to settle on a large scale by the middle of the 17th century [cf. Shepherd 1995]. Between 1895 and 1945, Taiwan was a Japanese colony. After World War II, the allied forces placed Taiwan under the sovereignty of the Republic of China which had been founded in 1912. When the Communists defeated the Nationalist troops in 1949, the nationalist government withdrew to China, which led to a large influx of immigrants from many parts of Mainland China.

Since 1945, the official language of Taiwan is Mandarin Chinese, generally referred to as *Guoyu* ('National Language'). About 70 percent of the population are native speakers of dialects belonging to the Southern Min group. These dialects are now collectively referred to as *Taiyu*, lit. 'the language(s) of Taiwan'. Although Taiyu is presently most widely used, the term has also been criticized as it suggests that Southern Min is the only local language of Taiwan, ignoring Hakka (*Kejia*) and the indigenous Austronesian languages. However, for stylistic reasons it seems reasonable to prefer the benefits of this short term to the political correctness of cumbersome alternatives. I will use the transliteration *Taiyu* instead of the English term *Taiwanese* only when unambiguous reference to Taiwanese as a linguistic variety is required.

1.2 Taiwanese orthography and lexicography

Since 1945, education in Taiwan has focused on spreading the use of Mandarin Chinese. Official treatment of local languages has since then shifted from systematic oppression before the 1980s, to toleration and, since the 1990s, modest inclusion in the school curriculum. The orthographic standardization of local languages has largely been neglected by official language planning agencies.

In sum, there is no authoritative norm for writing Taiwanese at the present. On the other hand, orthographic blueprints for the writing of Taiwanese abound. Since the late 1980s, these have been actively promoted by individual authors, language revivalists, local literary circles, as well as private research societies [Hsiau 2000: 125-147; Klöter 2001b]. Typologically, these proposals can roughly be divided into (a) solutions within the traditional Chinese character script and (b) alphabetical solutions, such as the traditional missionary system and innovative romanization systems. Within the employed writing systems, orthographic idiosyncrasy prevails.

Orthographic competition has left obvious traces in the 100-odd dictionaries for Taiwanese and related dialects produced over the last 150 years. The groundwork for an alphabetic writing of Taiyu was laid by Western missionaries. The first handwritten dictionaries of Southern Min dialects were compiled during the 16th century by Dominican priests and their translators working among the Chinese community in Manila [Masini 2000]. After the mid-19th century, protestant missionary lexicographers gradually developed an alphabetic orthography for Southern Min dialects known as Church Romanization [Heylen 2001]. For missionary lexicographers, Chinese characters were of secondary importance. Some missionaries even fully abandoned the Chinese script in favor of the Roman alphabet. Douglas [1873: viii-ix], for instance, regarded the "want of the Chinese character" as the "most serious defect", since "there are a very large number of the words for which we have not been able to find the corresponding character at all, perhaps a quarter or a third of the whole".

At present, the Maryknoll Language Center is the only missionary institution in Taiwan which continues to publish dictionaries and teaching manuals using the traditional missionary Romanization system [Maryknoll 1976, 1990-7, 1995]. During the late 1990s, however, alphabetic writing of Taiwanese has gained ground among language activists who argue that the writing system for the language spoken by the majority of the Taiwanese population should reflect a clear visible demarcation from Chinese writing traditions. The choice of a particular romanization system, however, remains the subject of fierce debates [Klöter 2001b]. The revival of alphabetic writing has thus far not resulted in significant lexicographic works.

The overwhelming majority of dictionary makers hold that only the use of Chinese characters can do justice to the genetic affiliation of Taiwanese with Mandarin. The choice for this writing system is not without its problems, as many expressions in regional Chinese languages lack established characters. According to an estimation by Cheng [1989: 332] "in Taiwanese, approximately 5% of the morphemes have no appropriate, established Chinese characters to represent them. Since many of these morphemes are high-frequency function

words, in a written Taiwanese text they account for as much as 15% of the total numbers of characters".

The dispute concerning the use of characters centers on the question whether written Taiwanese should reflect etymological correctness or a *couleur locale* by drawing from popular writing conventions. Popular conventions are attested in written editions of folk literature originating from Southern Fujian. The earliest written Southern Min source extant today is the stage play *Lijing ji* 'The lychee-mirror' (1566). One characteristic of the printed edition is its "consistent use of characters peculiar to Min-nan", i.e. to Southern Min [Van der Loon 1992: 3]. In a similar fashion, the local language of southern Fujian finds written expression in various genres of post-16th-century folk literature. These include manuscripts of lyrical and dramatic arias of the *nanguan*-opera societies, or ballad chapbooks known as *kua*^{1}-*a*^{2} *phoo*^{7}, the earliest editions dating back to 1826. For a short overview of the different genres of early Fujianese folk literature, I refer to Van der Loon [1992: 4-14].

Shi [2000], in his comprehensive study on the music and language of the *Lijing ji*, distinguishes three traditional techniques of rendering linguistic localisms:

1. The use of semantic loans, i.e. the adaptation of characters purely for their meaning regardless of the sound, as for example 在 (Mandarin *zài* 'be in') for Taiwanese *tu*^{6} 'be in' (p. 463);
2. Rebus writing, i.e. the adaptation of characters purely for their sound regardless of the meaning, e.g. in the use of 卜 for *boh*^{4} 'want, desire'; the original character represents the homonymous expression 'divine' (pp. 464f.);
3. The use of popular or dialect characters not attested in standard reference works, for example 煞 for *chua*^{7} 'lead the way' (p. 469).

Solutions in popular sources have served as a source of inspiration for some contemporary dictionary makers, among them Liu [1998], Wu [1998], and Qiu [1996, 1999].

From the viewpoint of traditional Chinese philology, however, it is especially the use of phonetic loans in popular sources which is unacceptable. A fixed etymology-based relation between a morpheme and its graphic representation is a matter of principle. Since the beginning of the 20th century, philologists have therefore attempted to counter popular convention with scholarly correctness. Their attempts are likewise reflected in a number of Taiyu dictionaries containing orthographies which preserve the correct etymologies of Taiwanese expressions. In these dictionaries, Taiwanese expressions are written with so-called *benzi* 'original characters', i.e. characters standing for assumed etymons in mainstream, i.e. Mandarin, literature.

In sum, the question of orthodoxy in the representation of Taiwanese expressions continues to be a bone of contention. As dictionary makers who cultivate solutions within the character script disagree on the principles of character selection, the outcomes of their efforts display a high degree of divergence. This divergence not only exists between the popular and the etymological factions, but also within factions.

2 Etymological orthography

2.1 Theoretical preliminaries

The distinction between etymological characters and loan characters has important theoretical implications. In linguistic theory, etymology as an orthographic principle has generally been discussed in connection with alphabetic writing, as for example by Coulmas [1999: 147]. He mentions the example of etymological spelling in English, briefly treating the examples of "silent" letters such as <l> in *folk*, <k> in *knife* and <w> in *wrestle* as etymological remnants. By contrast, the Chinese writing system is characterized as "logographic", for example by Sampson [1985]. In his words, "a graph of the Chinese writing system stands not for a unit of pronunciation but for a morpheme, a minimal meaningful unit of the Chinese language" [1985: 145]. This view, common in Chinese philology and Western sinology, is based on the notion of the Chinese writing system as representing a supradialectal standard language. Etymological origins of Mandarin morphemes and their graphic representations are attested in a rich literary tradition. On the other hand, the use of Chinese characters for the written representation of local Chinese languages has thus far hardly been considered in theoretical treatises of the Chinese writing system. The lack of attested links between linguistic localisms and early sources of Chinese writing weakens the etymological status of character usage in local literature. In this context, competing orthographic principles yielding different graphic representations of the same etymon need to be distinguished. The etymological principle may, in other words, be highly relevant to a non-alphabetical writing system.

2.2 Lian Yatang's *Dictionary of Taiwanese*

The first dictionary maker whose selection of entry characters followed strict etymological considerations was the Taiwanese historian and poet Lian Yatang (also known as Lian Heng, 1878-1936). When Lian embarked upon research on the etymological roots of Taiwanese in the late 1920s, he had already earned a reputation as one of Taiwan's most outstanding traditional literati. In 1957, Lian's etymological investigations were posthumously published in the dictionary *Taiwan Yudian* 'Dictionary of Taiwanese'.

As a traditional scholar, Lian sought the linguistic origins of Taiwanese expressions in the canon of scholarly literature. The list of works Lian quotes to provide evidence for his findings covers a wide range of genres from various periods. Lexicographic works include for instance the oldest Chinese character dictionary *Shuowen jiezi* 'Explanations of simple graphs and analyses of composite graphs' (AD 100) and the 11th century rhyme dictionary *Jiyun* 'Collected rhymes'. Other important sources are the Confucian Classics and poetry of the Tang Dynasty (AD 608 – 907).

Lian's reconstructions are based first and foremost on semantic associations. Questions of linguistic form are almost largely neglected, both as a criterion for reconstruction and in the presentation of entries. A typical entry in Lian's dictionary reads as follows (my translation):

> 查某 [*ca*^{7}-*boo*^{2} 'female']
> Women are called 查某 [*ca*^{7}-*boo*^{2}]. In the old days, women had a clan name, but no personal name. A woman is therefore called "someone", and for example referred to

as "someone's lady someone" or "someone's wife someone". Saying 查某 [*ca*^{7}-*boo*^{2}] is like saying 'this girl', as for example in the reference "this young lady" in the *Odes of Shao and the South* of the *Book of Poetry* [Lian 1957: 69; quotation translated by Legge 1871: 20].

2.3 Recent works

Endeavors to represent Taiwanese expressions with etymological characters have driven dictionary makers up to the present day. In 1998 and 1999, not less than three dictionaries were published, viz. Yang [1998], Mandarin Promotion Council (hereafter: MPC; 1998-9), and Zhang [1999]. A more recent publication, Dong [2001], largely follows the proposals in MPC [1998-9].

The MPC is the official language planning institution under the Ministry of Education. In 1995, the MPC initiated a research project to compile a four-volume lexicon for Taiwanese with approximately 600 monosyllabic entries represented by characters. Relevant philological research was primarily carried out by professor Yang Xiufang of National Taiwan University. Her drafts were examined and revised by two committees. By the spring of 2000, two volumes of the lexicon had been published [MPC 1998, 1999]. The publication of the final issues had been planned for 2001, but the project was temporarily halted due to budgetary problems.

The entries in the lexicon recommend spellings for the following three groups of expressions identified as problem areas:

1. Expressions commonly represented by idiosyncratic loan characters or dialect characters. *Siunn*^{1} 'too, exceedingly', for example, is frequently represented by the characters 尚, 嵩, or 爽 in popular writings. In the lexicon, 傷 is given as the original character for this expression.
2. Expressions that are used differently in Mandarin and Taiwanese even though they have clear semantic and formal similarities. For example, 25 sub-entries are listed under 頭 *thau*^{5} 'head', some of which have no corresponding etymon in contemporary Mandarin.
3. Formally similar Taiwanese expressions with slight semantic distinctions, to be reflected by different characters, e.g. 凭 *phing*^{7} 'lean one's body against' and 憑 *pin*^{5} 'rely on, depend on'.

The lexicon is based on the research of eminent scholars in the field. In terms of etymological correctness, it without doubt surpasses earlier dictionaries. Without laying any claim to completeness, the compilers have left aside some lexical groups, such as loanwords and expressions with obvious Mandarin cognates. Due to its narrow scope it is at best a first step to a standardization of a character script for Taiwanese. The tentative character of the MPC approach is not only reflected in the lexicon itself, but also in the absence of what Cooper has referred to as "essential features of the standardization of written languages", i.e. "written dictionaries, grammars, spellers, and style manuals" [1989: 145]. It is striking that none of the Taiyu teaching manuals currently available follows the spelling recommendations of the MPC.

2.4 Comparison of orthographic proposals

As stated above, the findings presented in the MPC lexicon are based on a much more refined methodological approach than Lian's pioneering investigations of the early 1930s. For detailed explanations of this approach, I refer to Yang [2000]. It is not surprising that different approaches have yielded different results. But has methodological progress led to a higher degree of orthographic uniformity? A comparison of three recently published dictionaries reveals that this is not the case. Orthographic divergence displayed by these works is immense, although all compilers claim etymological correctness. Table 1 compares the entries of three recent dictionaries [Yang 1998, MPC 1998-9, Zhang 1999] with Lian's [1957] findings. Characters separated by a slash mark (安呢 / 安爾) in line 1 are alternative spellings for the same etymon. Different characters for *bat*^{4} (example 2) stand for the two meanings 'know' and 'EXPERIENTIAL ASPECT MARKER' respectively.

	Taiwanese expression	English gloss	Lian [1957]	Yang [1998]	Zhang [1999]	MPC [1998-9]
1.	*an*^{1}*-ni*^{1} *an*^{1}*-ne*^{1}	'this, such, in this manner'	安仍	安呢/ 安爾	焉爾	--
2.	*bat*^{4}, *pat*^{4}	'know'; EXPERIENTIAL ASPECT MARKER	八	訊 剔	目 別	別 捌
3.	*be*^{7}	'not, not yet'	未	未	未	未
4.	*bueh*^{4}	'wish, just about to, will'	慏	爲	俾	--
5.	*co*^{3}*-hue*^{3}	'together'	做伙	做伙	湊伙	做夥
6.	*tai*^{7}*-ci*^{3}	'matter, affair'	戴志	戴誌	它事	事志

Table 1: Etymological spellings in four Taiwanese dictionaries compared

3 Concluding remarks and prospects

The differences in character selection exemplified in table 1 support one observation that will be confirmed by a broader comparison: orthographic solutions in Taiwanese dictionaries display a considerable amount of variants, even if the compilers are committed to the same principle of character selection. This high degree of diversity may be one reason why etymology has thus far had little influence on the orthographic development of Taiwanese. The etymologists, it is often argued, indulge in scholarly pedantry and ignore the popular heritage, which ultimately hampers orthographic standardization. For example, see the dispute between Hong [1988] and Cheng [1988].

Due to its narrow scope, the MPC lexicon will arguably contribute little to the reduction of orthographic idiosyncrasy. Official language planners are apparently aware of this modest influence. Hence, the Ministry of Education has recently initiated a new lexicographical project, aiming at the compilation of a Taiwanese dictionary for frequently used expressions, and comprising about 15,000 entries (cf. the project description at http://www.edu.tw/mandr). The compilation of a dictionary for the Hakka language has been initiated as well. The dictionaries are intended for primary and junior high school students.

Publication of printed and digital editions is scheduled for the year 2004. Unlike the MPC lexicon, the new Taiwanese dictionaries are obviously not intended as etymological correctives to orthographic idiosyncrasy. It will rather be attempted to weigh up etymological considerations against frequency of character usage in a variety of sources, including written sources of folk literature. Given the limitations of etymological verification as well as prevalent objections against an etymological orthography for Taiwanese, such an eclectic approach seems more promising than past attempts.

References

[Cheng 1989] Cheng, R.L., 1989. *Zou xiang biaozhunhua de Taiwanhua wen / Essays on written Taiwanese*. Zili Wanbao Wenhua Chubanbu, Taipei.

[Cheng 1988] Cheng, R.L., 1988. *Zheng Liangwei taolun* [R. Cheng's assessment], in: Cheng, R.L. & Huang Xuanfan (eds.) *Xiandai Taiwanhua yanjiu lunwen ji / The structure of Taiwanese: a modern synthesis*. Crane, Taipei.

[Cooper 1989] Cooper, R.L., 1989. *Language planning and social change*. Cambridge University Press, Cambridge.

[Coulmas 1999] Coulmas, F., 1999. *The Blackwell encyclopedia of writing systems*, Blackwell, Oxford.

[Dong 2000] Dong Zhongsi, 1996. *Taiwan Minnanyu cidian* [A dictionary of Southern Min of Taiwan]. Wu Nan Tushu Chuban Youxian Gongsi, Taipei.

[Douglas 1873] Douglas, C., 1873. *Chinese-English dictionary of the vernacular or spoken language of Amoy, With the principal variations of the Chang-chew and Chin-chew dialects*. Trübner & Co, London.

[Heylen 2001] Heylen, A., 2001. Missionary linguistics on Taiwan. Romanizing Taiwanese: Codification and standardization of dictionaries in Southern Min (1837-1923), in: W. Ku & K. De Ridder (eds.) *Authentic Chinese Christianity: Preludes to its developments (nineteenth and twentieth centuries)*. Leuven University Press, Leuven.

[Hong 1993] Hong Weiren (ed.), 1993. *Minnanyu jingdian cishu huibian* [An edition of classical reference books for Southern Min, in 10 vls.]. Wuling Chuban Youxian Gongsi, Taipei.

[Hong 1988] Hong Weiren, 1988. Tan Fulaoyu de zhengzi yu yuyuan [On correct characters and the etymology of Taiwanese], in: Cheng, R.L. & Huang Xuanfan (eds.) *Xiandai Taiwanhua yanjiu lunwen ji / The structure of Taiwanese: a modern synthesis*. Crane, Taipei.

[Hsiau 2000] Hsiau A., 2000. *Contemporary Taiwanese Cultural Nationalism*. Routledge, London and New York.

[Klöter 2001a] Klöter, H., 2001a. Recent lexicographical trends in Taiwan. Paper presented at the *Second Conference of the European Association of Chinese Linguistics*, University of Rome "La Sapienza", 6-8 September 2001.

[Klöter 2001b] Klöter, H., 2001b. Writing Taiwanese: the people behind the scripts. Paper presented at the workshop *Transformation! - Innovation? Taiwan in her cultural dimensions*. Ruhr University, Bochum, 7-9 March 2001.

[Legge 1871] Legge, J., 1871. *The Chinese Classics, Vol. IV, Part I: The She King or the book of poetry. With a translation, critical and exegetical notes, prolegomena, and copious indexes*. Henry Frowde, Oxford University Press, London.

[Lian 1957] Lian Yatang, 1957. *Taiwan yudian* [Dictionary of Taiwanese]. Zhonghua Congshu Weiyuanhui, Taipei.

[Liu 1998] Liu Zhenxiong, 1998. *Taiyu tongsu zidian* [A popular dictionary of Taiwanese]. Privately published by the author, Taipei.

[Maryknoll 1995] Maryknoll, 1995. *English-Amoy dictionary*. The Maryknoll Service Center, Taichung.

[Maryknoll 1990-1997] Maryknoll, 1990-1997. *Taiwanese*. Three volumes: Vol. 1 (1997), Vol. 2 (1990), Vol. 3 (1993). The Maryknoll Service Center, Taichung.

[Maryknoll 1976] Maryknoll, 1976. *Amoy-English dictionary.* The Maryknoll Service Center, Taichung.

[Masini 2000] Masini, F., 2000. Materiali lessicografici sulla lingua Chinese redatti dagli occidentali fra '500 e '600: I dialetti del Fujian [Lexicographical material of the Chinese language compiled by westerners in the 15th and 16th century: The dialect of Fujian], in: *Cina* 28, pp. 53-79.

[MPC 1999] Mandarin Promotion Council, Ministry of Education, 1999. *Minnanyu zihui (er)* [A lexicon of Southern Min, part II]. Taipei.

[MPC 1998] Mandarin Promotion Council, Ministry of Education, 1998. *Minnanyu zihui (yi)* [A lexicon of Southern Min, part I]. Taipei.

[Qiu 1999] Qiu Wenxi and Chen Xianguo, 1999. *Shiyong Taiwan yandian* [A Practical dictionary of Taiwanese sayings]. Zhangshu Chubanshe, Taipei.

[Qiu 1996] Qiu Wenxi and Chen Xianguo, 1996. *Shiyong Huayu Taiyu duizhao dian* [A practical comparative Mandarin - Taiwanese dictionary]. Zhangshu Chubanshe, Taipei.

[Sampson 1985] Sampson, G., 1985. *Writing systems: A linguistic introduction.* Stanford University Press, Stanford.

[Shepherd 1995] Shepher, J.R., 1995. *Statecraft and political economy on the Taiwan frontier.* SMC Publishing, Taipei.

[Shi 2000] Shi Binghua, 2000. *"Lijing ji" yinyue yu yuyan zhi yanjiu* [A research on the music and language of the *Lijing Ji*]. Wen Shi Zhe Chubanshe, Taipei.

[Van der Loon 1992] Van der Loon, P., 1992. *The classical theater and art song of south Fukien.* SMC Publishing, Taipei.

[Wu 1998] Wu Guoan, 1998. *Taiyu siyong hanzi ziyuan / The origins of hanji usage in Taiwanese: the advanced learner's dictionary for writing Taiwanese.* Privately published by the author, Taipei.

[Yang 1998] Yang Qingchu, 1998. *Tai-Hua shuangyu cidian* [A Taiwanese-Mandarin bilingual dictionary]. Dunli Chubanshe, Kaohsiung.

[Yang 2000] Yang Xiufang, 2000. Fangyan benzi yanjiu de guannian yu fangfa [Concepts and methods in the study of original characters of dialects], in: *Hanxue Yanjiu / Chinese Studies.* Special Edition No. 18 / Dec. 2000.

[Zhang 1999] Zhang Qingbo, 1999. *Taiyu zhengzi* [Correct characters of Taiwanese]. Tainan Xianli Wenhua Zhongxin, Xinying.