The Old Zhuang Script: a Systematic Overview

D.L. Holm
Department of Ethnology
National Chengchi University, Taipei
dvdholm37@gmail.com

Abstract:
The traditional Zhuang character script is an instance of a sinoxenic script, that is, a script in which the Chinese character script has been borrowed and modified to write a different language, in this case the Tai language now known as Zhuang.

In the south and southwest of China, the expansion of the Chinese state from the Qin dynasty onwards gave rise to a number of character scripts among non-Han Chinese peoples. The so-called Zhuang script is probably one of the oldest of these: reading pronunciations of individual graphs frequently correspond to Early Middle Chinese, and occasionally to even earlier strata of Chinese. With a history going back at least 1600 years, if not further, the script provides invaluable evidence about earlier stages in the Tai languages, and is of direct relevance for reconstructions of proto-Tai.

The Zhuang character script presents particular difficulties. Unlike some other sinoxenic scripts, the Zhuang script is a vernacular script rather than an orthographic script used at court. It is unstandardised and varies from place to place. The nature of this variation has never previously been documented. The aim of the author’s research has been to begin to address this issue systematically.

This paper will present the results of a survey of the script in 45 locations throughout Guangxi and contiguous provinces. Localities in Bouyei-speaking parts of Guizhou, Zhuang-speaking parts of Eastern Yunnan, and Sha, Nung and Tay-speaking parts of Northern Vietnam have been included. The aim is to provide a broad overview that encompasses basic regional and inter-regional affinities across the whole of the Zhuang-Bouyei-Tai area. The survey therefore transcends a number of administrative boundaries (including an international boundary) in order to document underlying linguistic and graphic continua.

Introduction

Let me begin with the observation that, broadly considered, the Tai peoples, in their present distribution, inhabit portions of both the Indosphere and the Sinosphere. The Indic and Chinese influences include writing systems but also many other aspects of language and culture. Although the Tai peoples inhabit what is still a single contiguous macroregion spanning mainland Southeast Asia and the southern part of China, the substantial continuities in language and culture have been masked until relatively recently by national boundaries, single-country scholarship, and a host of other factors. Comparatively speaking Thai Studies in Thailand is in a healthy state, and the results of decades of scholarly endeavours are reasonably well known and accessible internationally. The same cannot be said about research on the Tai peoples of Southern China: Zhuang, Bouyei, Kam, Sui, Gelao, Hlai, and so on.

One of the major stumbling blocks has been that the traditional written literatures of these people present particular difficulties, and have only recently become available (but in Chinese editions, not in English or Thai). A systematic approach to the Chinese-style character script in which these literatures are written is badly needed.

Previous Approaches

Previous studies of the Old Zhuang Script have concentrated on the script of particular localities, and mainly on the combination graphs invented to represent Zhuang morphemes. It is entirely
natural that these interesting graphic phenomena should excite interest. Our Chairperson Bob Bauer has himself contributed to this field in an article published in Cahiers LAO in 2000.¹ Most recently, in 2010, Qin Xiaohang at the Central Nationalities University in Beijing has produced a monograph on the subject.² This too is based on the analysis of isolated graphs, many of them taken from the 1989 dictionary Sawndip sawdenj (Dictionary of the Old Zhuang Script).

Unfortunately this is not the best way to start. The Sawndip sawdenj is a useful compendium, but it provides no information about where the dialect forms come from, so it is impossible to see any patterns in geographic variation from this source. In any case the special Zhuang characters need to be contextualised as part of a wider system before they can be properly analysed. It is to address these wider systematic questions, and to enquire more carefully into such matters as the age of the script and the nature of its geographic variation, that I undertook this study on which I am reporting.

The Handbook

My findings are written up in a monograph which I am calling A Handbook of the Old Zhuang Script. I am not used to blowing my own trumpet but in this case I can be reasonably confident that this will become a milestone in the field of Thai Studies as well as in Sinology. I am assured by Ting Pang-hsin and other eminent scholars in the field that this is the first time that any of the Sinoxenic character scripts of South China has been analysed systematically. By systematically, I mean that the whole area in which the script is current has been surveyed, texts from each major area have been analysed, and the results have been compared cross-regionally both script- internally and in relation to dialect differences on the ground, pre-modern readings of the relevant Chinese characters going back to the Han dynasty, dialect readings of Chinese characters, and proto-Tai and proto-Kam-Tai reconstructions. The result is a very substantial work of reference that is intended at the same time to serve as a demonstration of method and a springboard to future research.

For the benefit of our Thai colleagues I should explain. Unlike the Tai languages further to the west and southwest, Tai languages within China and in the northern part of Vietnam were (and are) written in a character script that was based on the Chinese script. Unlike Indic scripts, the units of which relate variously to the consonants, vowels and tones that constitute syllables, the Chinese script is a square-block script with units representing mono-syllabic morphemes. Many characters in Chinese have more than one reading pronunciation, but many have only one. Most characters, when used to write Chinese, contain a semantic component (a so-called radical or semantophore), and a phonetic component (phonetic or phonophore). When borrowed to write Tai or other languages, however, the graphs were taken over holus-bolus. Scribes were faced with a choice between using a Chinese character with the same meaning (a semantic borrowing) and using a Chinese character with the same sound (a phonetic borrowing). In the case of the Tai languages ancestral to Zhuang (Bouyei, Nung, Tay, Sha, and so forth), the process was facilitated by the fact that Chinese and Tai were typologically similar and had inventories of initial consonants, vowels, codas and tones which were not entirely dissimilar, at least had a substantial area of overlap.

So far so good. You would think, under such circumstances, that reading Tai texts in this script would still be reasonably straightforward, since you were faced with two basic choices for each

graph, and with a little trial and error you should be able to make sense of things. Here you would be wrong. There are in fact far more than just two possible readings for each graph (I’ll come back to this in a minute), and with the phonetic readings we find ourselves faced with a vast range of possibilities. The Tai languages in China were in close association with Chinese over a vastly long period, from the Chinese conquest of the South in 221 BCE right up to the present. That’s a period of 2200 years. The period when actual borrowings took place, resulting in reading pronunciations in the script, stretches between the Han dynasty and the Ming and Qing dynasties, in other words between 100 BCE and around 1800. I am leaving out of account here the recent influx of modern Mandarin since 1949, and Cantonese since 1800 or so. That’s still a period of some 1900 years. What our investigations demonstrate is: Reading pronunciations of Chinese characters in Tai texts either become fixed at some point (and conventionalised within the Tai script system), or continue to change as the local pronunciation of Chinese and the Tai language change. So in effect, what you have is a layer cake, with the oldest stratum of readings going back to the Han or thereabouts, and the most recent reflecting Southwestern Mandarin readings that became the lingua franca from the beginning of the Ming dynasty (1368-) onward.

So what the Handbook does is go systematically through all of these possibilities for a set of 60 common words. All of the different ways that these 60 common words are represented in the texts surveyed have been collated and analysed. The survey is based directly on traditional texts from 45 different localities (map on PPT), and this is sufficient to give us a very large amount of information about the script: some 3000 different graphs and 689 graphic series. Also it demonstrates how the script differs from locality to locality, and also how the different local varieties of the script are related to each other. Armed with the Handbook, it will be possible for scholars with a reasonable level of familiarity to look at a traditional Zhuang manuscript and say where it comes from. (example hawj ‘to give’ on PPT)

I mentioned above that there are more than just two possible readings of Chinese-style characters in traditional Tai texts. In fact there are 12 (so far). These are:

**Straight borrowings:**

phonetic borrowings (examples on PPT)
semantic borrowings
phonetic/semantic approximates

**Serial borrowing:**

semantic + phonetic (homophone substitution)
phonetic + semantic (synonym substitution)
semantic + semantic
phonetic + semantic + phonetic

**Catalytic readings:**

part for whole
simplex for compound
reconstituted compounds
borrowed graphs (graphic approximants)
modified phonetic readings
This typology comes directly from a close analysis of the way a body of texts are recited by their traditional owners (Holm 2003 and 2004). It is important to point out here, methodologically, that you cannot get at this kind of information if you just stick to looking up isolated words in dictionaries.

Findings

Below I summarise some of my major findings thus far.

Regional Systems

A major finding of this study is that the graphic representation of each Zhuang morpheme in the survey falls into relatively clear regional patterns, and that the patterns of distribution for the 60 common words often evince a degree of congruence. Quite often, there are at least three regions: (slide on PPT)

* The Central Guangxi Region
  comprising Liuzhou, Laibin, Xincheng, Shanglin and Mashan
* The Youjiang and Zuojiang Region
  comprising Guiping, Wuming, Tiandong, Tianyang, Bose, Jingxi, and Napo in Guangxi, and Cao Băng and Lang Sơn in Vietnam
* The Northwestern Region
  comprising Yishan, Huanjiang, Donglan, Tianlin, and Tian’e in Guangxi; Libo, Ceheng, and Zhenning in Guizhou; and Xichou in eastern Yunnan

In other cases, the following regional systems can be discerned:

1. The West River valley in Central Guangxi
2. East-Central Guangxi
3. The Southwest and northern Vietnam
4. Tianzhou
5. The North and Central North, based on the Longjiang River valley in northern Guangxi;
6. The Northwest, Guizhou, and Eastern Yunnan

The regional pattern was found to vary somewhat from morpheme to morpheme, in much the same way as the geographic boundaries between pronunciations of dialect words is found to vary from lexeme to lexeme.

For many of the common words in the survey, there may actually be only two or three major representations. That is to say, Tianzhou, for instance, or the Southwest, may be the same as the West River valley system and central Guangxi, or the North and Central North may be the same as Guizhou. The lines of division between different graphic realisations are not entirely congruent – or not predictably so – but then isoglosses in dialectology do not line up in any simplistic way, either. Moreover, these systems are not watertight, and there is a lot of mutual interaction evident in the data. On the other hand, within the sub-regions there is a much higher incidence of common representations of common morphemes.

These things can now be demonstrated statistically. If we collate all the ways in which each of the 60 common words are represented in each locality, the degree of commonality can be calculated
in various ways. A simple first-approximation measure is to compare only the most common graphs used in each locality, on the basis of whether or not they are found within the same graphic series (counting 1 for same and 0 for different). (maps on PPT)

**Graphic Sub-systems**

Another major finding is that the regional systems vary in graphic composition. Very briefly:

1) the central Guangxi and northwestern systems are characterised by graphic conservatism. That is, most of the graphs preserve their original Chinese form, and any additional made-up Zhuang characters are based closely on the conventional Chinese rules of character composition – the so-called “Six Ways of Writing”. So the semantic components are almost always graphs which are recognised as among the conventional 214 “radicals”. Even characters which are read “catalytically” take the form of orthographic Chinese characters.

2) East-Central Guangxi is characterised by graphic innovation through radical re-analysis and simplification of commonly used graphs. Many characters are found in these scripts which are not found in Chinese.

3) In the opposite direction, the SW and northern Vietnam are characterised by graphic innovation in the direction of greater graphic complexity. Additional Zhuang characters have semantic components which are not confined to the 214 “radicals”, but more directly represent the semantic field of the word in the local language (“semantic indicators”). Many compound graphs have two complete Chinese characters side by side or one on top of the other, and have roughly twice as many brush-strokes as conventional Chinese characters.

**Script and Dialect**

The script varies in its relation to the local dialect pronunciation. In some localities, it can be shown that local pronunciation has had a direct influence on its graphic representation in the script. In other cases, graphic representation has a more tenuous relationship to local pronunciation. In these latter cases, there are various possibilities that need to be investigated locally. In many cases a particular way of writing Zhuang has been imported from a different locality, and the graph in question represents the local pronunciation in the source area. Another possibility is that the graph became fixed in form quite some time ago, and the local pronunciation has changed. In either case, we are dealing here with a writing system which does not directly reflect current dialect distributions.

The presence of graphs imported from other areas means that texts bear traces of their own history. The implications of this are very exciting. We can often tell from internal evidence that either texts of priestly lineages originated at least in part somewhere else. Comparison of scripts cross-regionally, as well as on-site interviewing, sometimes allow us to pinpoint where those places of origin were, and more often give us a general idea about the regions or dialect areas from which the texts or parts of texts came from. In many cases this alerts us to previously undocumented cultural flows.

**Pinghua and Southwestern Mandarin**

The phonetic basis of these systems exhibits a more clear-cut distinction, between systems in central Guangxi, the SW and Vietnam, which by and large are based on Pinghua readings of
Chinese characters, and the systems of the North, Central North and Northwest of Guangxi, plus Guizhou, which are based on Southwestern Mandarin readings of Chinese characters. Eastern Yunnan seems a mixed system, with some Pinghua readings but also a very strong overlay of Southwestern Mandarin.

A first approximation estimate of the relative age of the systems is based on this division between Pinghua and Southwestern Mandarin (SWM). Pinghua readings may be as early as the beginning of the Tang dynasty or even earlier. However, for some of the graphic systems in the present survey, historic sound change in the Tai languages and in Pinghua have running roughly in parallel, so that on the basis of the data we have reviewed (mindful of its limitations), the best we can do in many cases is say that the pronunciation on which the Zhuang character is based could have been borrowed at any time in the last 1700 years. This is an effect of languages in close contact.

On the other hand, the systems based on Southwestern Mandarin could only have taken shape after the beginning of the Ming dynasty, after the military takeover of Guizhou by the Ming armies starting around 1390. These systems have a maximum age, therefore, of only around 600 years – less than half the potential age of the script in central Guangxi.

**Proto-Tai**

The approach outlined here promises to open up a very considerable body of new data for the reconstruction of Proto-Tai. The oldest strata of readings are at least as old as the Qieyun 切韻 system (601 CE) and some may be considerably older. I have drawn attention in the Phonology section of each chapter in the Handbook to these correspondences and also to cases in which Middle Chinese, Early Middle Chinese and Late Han readings are in close correspondence with Proto-Tai and Proto-Kam-Tai reconstructions. Potentially at least the OZS data gives us information about the Tai languages at twice the historical time depth of the King Ramkanghaeng inscription of 1283 – i.e. 1400 as opposed to 700 years BP. This is a very considerable difference, and potentially takes us back to an era very close to the time when, according to a recent study, the exodus of the Tai people from southern China began and the split between the SW, CT and NT languages took its present form.³

This has an important implication: It means that it is likely that the ancestors of the Thai people used a character script, based on Chinese and in some sense ancestral to the old Zhuang script of the present day, before their exodus from Southern China and migration to the west and southwest, where they came into contact with Indic scripts.

Li Fang Kuei’s Proto-Tai reconstructions will need to be re-evaluated in light of this new data, just as Liang Min and Zhang Junru’s reconstructions of Proto-Kam-Tai need to be re-evaluated on the basis of better data for Tai languages outside China.

Of course, 60 words is too narrow a base from which to launch a major re-evaluation of Proto-Tai. Nevertheless, I should offer some observations on how this might be done in future, incorporating consideration of the OZS data.

³ Huang Xingqiu 黄兴球, *Zhuang-Tai zuqun fenhuaxianzhong kào* 壮泰族群分化时间考 [An Investigation into the date of the split between the Zhuang and Thai peoples], Beijing: Minzu chubanshe, 2008. Using a variety of independently variable linguistic and cultural indicators. Dr. Huang dates the beginning of the split to the 6th century CE.
This survey has shown that the oldest strata of readings are found in central and east-central Guangxi, though the EC system represents graphically a fairly recent development. OZS further east in counties that used to be solidly Zhuang-speaking may be even older. Texts from these eastern areas should be edited and locality-specific dictionaries produced. These could then be used as a broad basis from which to assemble graphic representations of an appropriate base vocabulary for PT reconstruction. Comparison of the phonetic components of these graphic representations could then take place with existing PT and PKT reconstructions, along the lines I have undertaken here, and appropriate proposals made. It is important that this is done systematically, for all proposed PT and PKT initials and finals. This would also enable us to take into account and explain more adequately the reflections in the OZS of words for which the Chinese script did not provide adequate phonetic matches. The eventual result, I expect, will be a re-confirmation of some of Li Fang Kuei’s reconstructions, revision of others, and possibly some major re-thinking of old problems.

**Innovation and Integration**

The existence of the regional systems as demonstrated in this survey points to the operation of two important cultural mechanisms at work in the Tai-speaking areas of South China over many centuries: broad regional integration, and local innovation.

Local innovation is to be expected and is easy to understand. After all, one of the names for the traditional Zhuang script is *sawndip* ‘uncooked characters’. This term refers to the circumstance that the script was variable, and contained impromptu renderings that were made up on the spot. It has long been recognised that one of the functions of the Zhuang character script was to provide a rough-and-ready means of writing the Zhuang language using Chinese characters. At its most individual this mechanism amounts to private note taking, with an intended readership confined to one person – the writer. The widespread availability of primary education in the Zhuang and Bouyei speaking areas since the 18th century has provided a relatively broad basis for this practice.

The existence of broad regional integration in the Zhuang script is more surprising – or at the very least, it requires explanation. The Zhuang, unlike the early northern Chinese, did not have any centralised political authority under their own control that could have instigated a policy of unifying such a script. Such political authority as was in the hands of the Zhuang and Bouyei – native chieftaincy – was geographically fragmented for the most part, and only rarely on a scale resembling the broad graphic systems I have delineated. Zhuang village society is usually seen as localised and small-scale, though this also is a general conception that may be in need of revision.

One of the major factors contributing to the observable pattern of broad regional integration in the vernacular script was the influence of religious practitioners. In central and east-central Guangxi, the spread of organised groups of ritual masters of the Meishan Taoist tradition in the Ming and Qing was probably one of the main factors leading to the formation of the east-central graphic system. Ritual masters were also active in the upper reaches of the Longjiang system, and their

---


5 On this problem, see Holm, ‘Evidence for Historical Sound Change in Traditional Zhuang Texts’, 2010. There is considerably overlap between Zhuang-Tai initials and finals and Chinese of various periods, but also some initials and finals not found in Chinese.

influence extended as far as Donglan, Du’an, and Libo in southern Guizhou. Further west, Maoshan Taoists and mogong were similarly active. At a different level, the circulation of written song texts and chapbooks of other kinds of ceremonial songs, common throughout Guangxi and contiguous regions, was probably an important factor. After the Ming and Qing, the rise of storytelling forms, song-and-dance, and forms of theatre and marionette theatre in the Zhuang and Bouyei speaking areas provided an additional medium for region-wide circulation of texts.

**Future Prospects**

The project I have just outlined is beyond the capacities of any one individual. I would be very pleased to welcome collaboration. There is much work for young scholars with an interest in this field.

Absolutely essential is Fieldwork: close collaboration with traditional owners of texts, high-quality audio recordings of recitations, interviews on a line-by-line and morpheme-by-morpheme basis, high-resolution photographs of manuscripts, all described in Holm 2003.

Other skills that I have found useful:
- knowledge of Zhuang and other Tai languages
- knowledge of Chinese
- some understanding of Chinese historical phonology
- knowledge of Taoist and/or Buddhist religious traditions

**Recent Works by the Author**


4) ‘Evidence for Historical Sound Change in Traditional Zhuang Texts’, Journal of Language and Linguistics (Thammasat University), 28:2 (Jan-June 2010), 1-12

5) ‘Linguistic Diversity along the China-Vietnam Border’, Linguistics of the Tibeto-Burman Area, 33.2 (October 2010), 1-63

6) ‘Chuantong fangkuai Zhuangzi de quyuxing’ 傳統方塊壯字的區域性Regional variation in the traditional Zhuang character script], Guangxi minzu daxue xuebao 廣西民族大學學報, 2011 No.2, 27-33. (An English-language version of this article is available on request.)


---

7 D. Holm, *Killing a Buffalo for the Ancestors*, De Kalb: Southeast Asia Publications, Northern Illinois University, 2003, pp. 173-175, documents the out-migration of Zhuang-speaking Maonan ritual masters to the counties surrounding Huanjiang.