Evidence Regarding Decomposition of the VA Mark in Tai Lue Script

L2/04-272

Date:	2004-06-015
Author:	Peter Constable, Microsoft
Address:	One Microsoft Way
	Redmond, WA 98052
	USA
Tel:	+1 425 722 1867
Email:	petercon@microsoft.com

1   Introduction

Document N2634 presented a proposal for encoding the Tai Lue script, which proposal was by WG2 in M44 (resolution M44.1.e). Subsequently, in ballot comments on the PDAM for Amendment 1 (cf. SC 2/N3730) and in the Chinese contribution to WG2, N2748, China requested various revisions to what had been approved at M44. Among these was a request that certain characters representing combinations of a consonant with a labialization 'VA' mark be decomposed; corresponding to this, the proposed revision included the addition of a new character for the combining 'VA' mark.

The Chinese request in N2748 did not provide specific argumentation to support the request for encoding a combining 'VA' mark. This document, therefore, considers evidence that would be relevant in making this decision.

2   Some basic criteria

It is assumed here that certain basic criteria exist for determining when to encode a separate combining mark that can dynamically compose with base characters rather than encoding atomic characters for particular base + diacritic combinations. For instance, one criterion that is assumed is that a diacritic mark must have some identity independent of any particular base character in order for it to be encoded as a separate character, and likewise that the base must have some identify independent of the diacritic.

Another criterion assumed to be necessary is that the mark in question should be able to combine productively with various base characters. If a mark occurs only on one base character, or a very small number (e.g. two, three) of base characters, then there is no particular economy gained by encoding it as a distinct combining character rather than encoding the particular base/diacritic combinations as atomic characters.

These two criteria are considered necessary for encoding a combining mark. It is important to note, on the other hand, that they are not necessarily assumed here to be *sufficient* criteria.

These criteria, and in particular the latter, are relevant for considering the encoding of the VA mark. Treating the VA mark in Tai Lue as a combining mark introduces a potentially large degree of complexity to the implementation of the script. To justify encoding it as a combining mark, therefore, we should expect to find a number of dynamic combinations in which the VA mark would be used. If it is only used in a small number of combinations, however, the overall economy of the encoding implementation may be improved by representing each combination as an atomic character.

3 Evidence that might be used against encoding a combining VA mark

There are various descriptions of the Tai Lue script from which samples have been made available—linguistic works, literacy books and dictionaries—and in all of these, there are four labialized consonants listed (i.e., four consonants with the VA mark): high and low /kwa/ and /xwa/.

Figure 1. Four labialized (velar) consonants listed in SCEM, p. 72.

ဘွခလးၿၡင်မၣ

ၵသၡသၟၡ   ၵသၡၵၵၰမၞၔ

ဘဲ   ဘံ့

ဢ ၑ ၮ   ၵ ၆ ၡ

ၕ ၒ ၸ   ၡ ၦ ဃ

ၕ ၡ ၮ   ၦ ၕ ၣ

ၸ ဠ ၮ   ၮ ၒ ၿ

ၸ ၮ ၮ   ၥ ၀ ၼ

ၱ ၳ ၀   ၢ ၣ ၀

ၮ ၡ   ၰ ၦ

Figure 2. Four labialized (velar) consonants listed in LA1, p. 72.

This corresponds to the linguistic facts of the Tai family of languages: in reconstructions of proto-Tai phonology, there are only labialized variants for the velar consonants. (See, for instance, F-K Li.)

Beyond the historical Tai roots, languages of the region often have loans from Pali or Sanskrit origins, and loans of Indo-Aryan origin can involve many different clusters of the form /Cv/. Yet these other combinations are not cited as involving a VA mark in the various sources that have been available. Rather, it has been reported to me (William Hanna, personal communication), that these are written with the full letter VA. For instance, "၆ဝၱၵ" /tawiːp/ 'continent'. Reportedly, then, the VA mark is not used in these combinations.

In addition to the four labialized velar consonants, of course, there is also the high class VA, which is written like HA with the VA mark, "ၮ", and also the abbreviation for the particle /lew/, "ၮ". Beyond these two special cases, though, these documents show no productive use of the VA mark except with the velar consonants.

To preserve the overall economy of the encoding, therefore, it could be argued that it is best to represent each of these combinations as pre-composed, atomic characters.

4   Evidence that might be used in favour of encoding a combining VA mark

What is reported above comes from certain sources. Other sources offer differing evidence, however. In particular, I received correspondence from Mr. Yin Jianmin (of Huaguang Technology Co.) indicating that other combinations for labialized consonants were, in fact used:

high and low /sva/ and /tva/, and low /lva/. He provided samples taken from documents comparing the new Tai Lue script with old Tai Lue writing, derived from Lanna script.

So, for instance, in Figure 3, the text in the blue highlight is in the old orthography (which can be distinguished by the sub-joined NGA and the diacritic mark for /o/), while the text in the red highlight is in the newer Tai-Lue-script orthography:
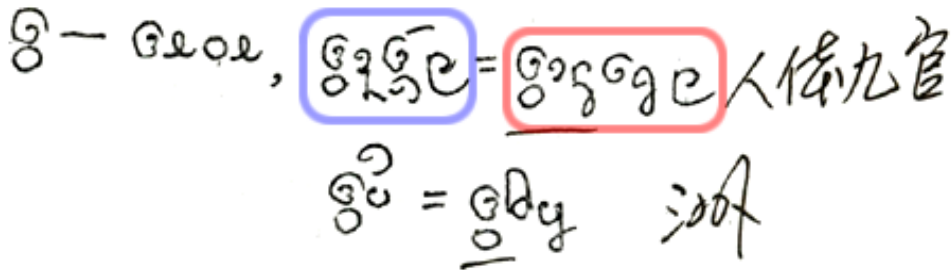

Figure 3. Libialized low /ta/ in comparison of old and New Tai Lue writing (Dao 1989)

The portion of the sample using the Tai Lue script clearly shows the VA mark combining with the low TA character.

The set of samples provided by Yin Jianmin showing the VA mark on consonants other than the four velar consonants are shown in Figure 4 and Figure 5:
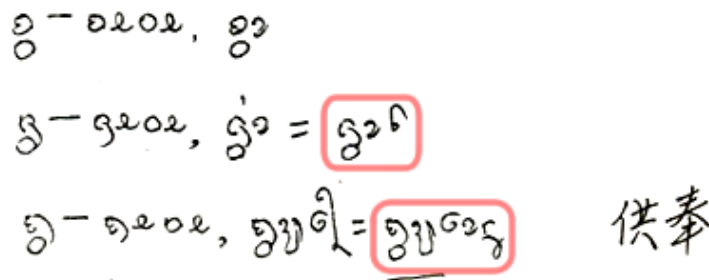

Figure 4. Libialized low /cwa/, high /thwa/ (Dao 1989)

Figure 5. Libialized high /thwa/, low /twa/, low /lwa/, high /swa/ and low /swa/ (Dao 1989)

In these examples: there are several consonants in addition to the velar consonants that take the VA mark: high and low /sa/, low /ta/, high /thaw/, and low /lwa/.

What is particularly interesting to note in these examples is that they involve comparisons of the old and new writing systems for Tai Lue script. This appears to suggest that there may be some variation in spelling conventions, with some authors writing some clusters using full VA, but with others using the VA mark on various consonants, under the influence of the old script, as shown here.

The issue here is that it suggests a potential for a large number of libalized combinations that occur in Lanna script, and presumably, potentially in the Tai Lue script as well. For instance, for Northern Thai text, the following clusters (high and/or low versions) are written in Lanna script

using a sub-joined VA mark: /kw/, /kʰw/, /tw/, /ʔw/, /cw/, /sw/, /jw/, /ŋw/, /ñw/ and /lw/. This suggests, then, a potential of twenty different labialized consonants. One might argue that to encode all of these as atomic characters would be less economical than encoding the VA mark as a separate character.

Of course, this argument is based on hypothetical combinations that *could* potentially occur in Tai Lue script, but that haven't been attested yet. There is another problem, though: as the examples in Figure 3, Figure 4, and Figure 5 are all taken from a context of comparing the old and the new script, it is natural to wonder whether the author was writing Tai Lue script as it is conventionally defined, or whether the author was intermixing scripts. The fact that sources *other than* such comparisons list only the velar consonants could perhaps be taken as indicating that these examples involve intermixing of the old and new scripts.

5   Conclusion

The evidence regarding productive combinations involving the VA mark is not entirely clear. Several references have been seen supporting only velar consonants in labialized form, while one source has shown several other consonants having velarized forms. On the other hand, that single source may not be representative of conventional use of Tai Lue script. The status of the samples provided by Yin Jianmin needs to be determined in order to clarify whether they constitute valid evidence for a separate VA mark, or whether they constitute nothing more than irrelevant cases of mixing scripts.

6   References

[SCEM]   中国社会科学院民族研究所，国家民族事务委员会文化宣传司 (Zhongguo shehui kexueyuan minzu yanjiusuo, guojia minzu shiwu weiyuanhui wenhua xuanchuansi = Chinese Academy of Social Sciences, Institute for Nationality Studies; Department of Cultural Propagation, State Ethnic Affairs Commission). 1991. 中国少数民族文字 (*Zhongguo shaoshu minzu wenzi = Scripts of Chinese Ethnic Minorities*). 北京 (Beijing): 中国藏学出版社 (Zhongguo zangxue chubanshe = China Tibetan Studies Publishing House).

[LA1]   西双版纳州教育局编 (Xishuangbanna Prefectural Education Bureau). 1980. ᦞᦵᦵᦲᦱᦺᦟᦺᦲᦵᦲᦱ, ᦵᦶᦶ ᦲᦲ ᦵ (语 文, 第一册 = Language arts, book 1). (西双版纳傣文小学课本 = Xishuangbanna Dai primary school textbook.) 昆明 (Kunming):云南民族出版社出版 (Yunnan Nationalities Press).

[F-K Li]   Fang Kuei Li. 1977. *A handbook of comparative Tai*. Honolulu: University of Hawaii Press.

[Dao]   Dao Xinhua. 1989. *Comparison Textbook of Old and New Tai Lue*.