## ISO/IEC JTC1/SC2/WG2/IRGN2434

Universal Multiple-Octet Coded Character Set International Organization for Standardization

Doc Type: ISO/IEC JTC1/SC2/WG2/IRG
Title: Request to dis-unify U+722B

Source: Vietnam

Status: National Body Proposal

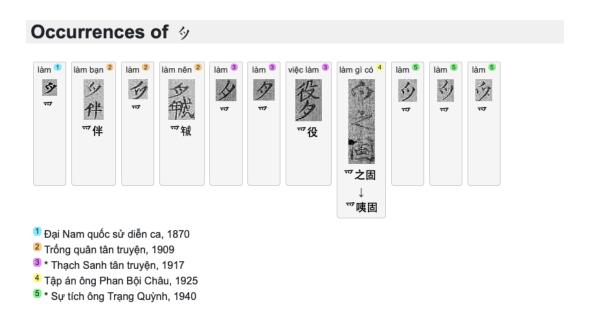
Authors: Eiso Chan, Lee Collins, Ngô Thanh Nhàn

Action: For consideration by the IRG

Date: 2020-06-10

This paper proposes dis-unification of URO character U+722B  $\overrightarrow{w}$  into 2 characters: The original U+722B, which would retain it shape and semantic as Kang Xi radical no. 87, and a new character designed to be centered within the ideographic bounding box, and to occupy the greater part of the bounding box space. The new character will be semantically different from U+722B, representing the Vietnamese verb  $l\grave{a}m$  and other words written with a full form of U+722B, such as Zhuang lamz.

Làm, meaning "make, do, consist of", etc. is one of the most frequently used Vietnamese verbs. It has many variant forms in chữ Nôm, including: 溢, 冷, 溢, and 今, but the most common forms are those shown in the picture below, with examples from various sources, as documented at <a href="https://www.chunom.org/pages/722B/">https://www.chunom.org/pages/722B/</a>. Detailed images are shown in the Appendix.



There are two source characters from which this form of  $l\grave{a}m$  appears to have been derived. One is U+7232  $\ \ \$   $\ \ \$  . The semantic of  $l\grave{a}m$  is clearly related to the sense of U+7232, but the shape has been reduced to the top component of U+7232. The resulting simplification facilitates writing of this high frequency character and marks the usage as Vietnamese. The second is the character U+6FEB  $\ \ \ \$  which has the Sino-Vietnamese reading lam. This derivation is purely phonetic. Nguyễn Quang Hồng's  $\ \ \$   $\ \ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \ \$   $\ \ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \$   $\ \ \$   $\ \$   $\ \ \$   $\ \$   $\ \ \$   $\ \ \$   $\ \ \$   $\ \ \ \$   $\ \ \$   $\ \ \$   $\ \ \ \$   $\ \ \ \ \$   $\ \ \ \$   $\ \ \ \$   $\ \ \ \ \$   $\ \$ 

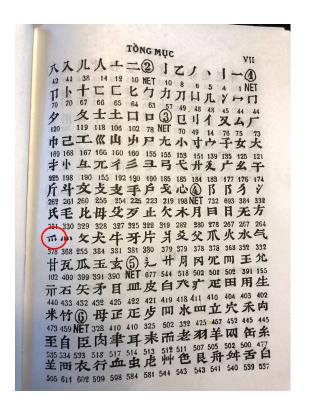
It is also possible that there is some kind of borrowing relationship between Vietnamese and Zhuang.

The  $\checkmark$  variant of  $L\grave{a}m$  was first encoded in TCVN 5773:1993 as V0-3D5B. When chữ Nôm was standardized in Unicode, V0-3D5B was unified with U+722B on the basis of the first source derivation described above.

Based on our experience with the use of U+722B, this was a mistake.

U+722B is properly a variant of U+722A  $\mathcal{M}$ , read  $tr\mathring{ao}$ , meaning claw, talon, etc. It is used as the component form of Kangxi radical 87, and not cognate with the sense of  $l\grave{a}m$ . The problem arising from the lack of cognateness manifests itself in several ways. One is that Vietnamese also requires U+722B as a the component form of  $tr\mathring{ao}$  for use in radical look up tables, illustrating

character formation, etc. This is the proper use of U+722B and the form used in these cases matches the shape of U+722B as drawn in most CJK fonts, as illustrated below.



Làm, should be drawn to occupy the full ideographic character box, so as to be in proportion with other Nôm characters. There are many situations where it is impossible to represent both forms without resorting styled text. And, since virtually every CJK font implements the component form of U+722B, where an application uses text that falls back to the system font, which is universally CJK, not Vietnamese, *làm* is rendered incorrectly with the component form of *trảo*. Below is an example from the Mac OS, where *làm* is rendered correctly in the document, but appears as the *trảo* component in the title.



Our proposal then is to add a new character representing *làm* in one of the CJK Unified Ideograph Ranges. It would have the same properties as U+722B, but the Unihan data would reflect the new semantics

## U+xxxxx

kIRG VSource V0-3D5B

kVietnamese làm kRSKangXi 87.0 kRSUnicode 4.3 kTotalStrokes 4

The V-Source glyph for U+722B would change to resemble the component forms shown above for the other sources. Also, two fields of the data for U+722B would likewise change:

U+722B

kIRG VSource VU-0722B

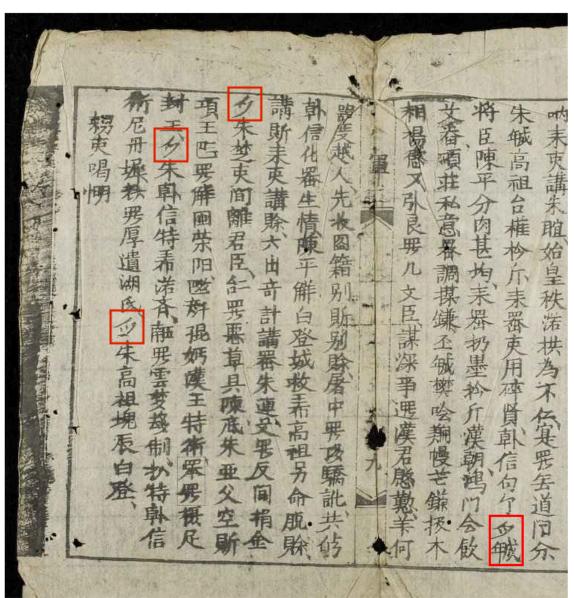
kVietnamese trảo

If China decides to standardize the Zhuang character *lamz*, they may also want to consider unification with the new code point. Although the semantics are different, there does appear to a relationship of borrowing or common derivation, that would justify unification. WS2017-03573:USAT05603, is a different matter. Although it resembles some of the forms of *làm*, the IRG might not consider this to be a valid candidate for unification based on the semantic. WS2017-03573:USAT05603 is a variant of the character, 肉 *nhục* (flesh), and has no relationship other than the similarity of some forms.

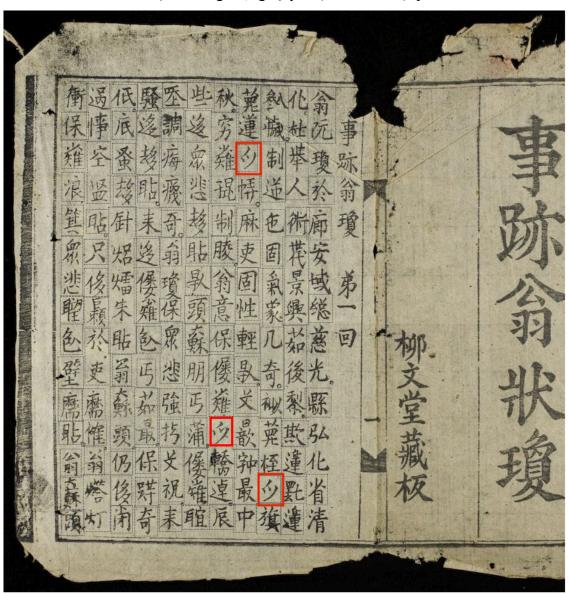
## 1. Đại Nam quốc sử diễn ca (大南國史演歌), p1



舸 国 吏 会 立 詞 姉 固 スニ 立 娟 謀 謀 講 雱 拢 国 媽 衛皇 姬 騑 7 船 婵 刻 竜 玄 巫 跳 於 甑 进 船 蠅 操 砌 如 脏 雷 图 伯 将 逷 志高、



6. Sự tích ông Trạng Quýnh (事跡翁狀瓊), p.1



## Bibliography

Guangxi Zhuangzu Zizhiqu Shaoshu Minzu Guji Zhengli Chuban Guihua Lingdao Xiaozu (ed.). *Sawndip Sawloih* (古壮字字典). Guangxi Minzu Chubanshe. Nanning, 2012.

Nguyễn Quang Hông. *Tự Điển Chữ Nôm Dẫn Giải*. [Dictionary of Chữ Nôm with source citations]. Social Sciences Publishing House, Hà Nội, 2014.