

Title: Proposal to add kZhuang to Unihan
 Author: John Knightley
 Date:15-03-2024

Zhuang characters have since CJK Extension F become an established and respected part of the Unicode community, and Unihan has accepted kZhuangNumeric therefore it is proposed that kZhuang also be added to Unihan. To date 2,471 Zhuang characters have been encoded having a GZ- source from the *Ancient Zhuang Character Dictionary* (古壮字字典), 1989, ISBN 7-5363-0614-8 and this can provide suitable guidance for a format for kZhuang. All readings in the dictionary are written using the official Zhuang orthography which uses the 26 letters of the English alphabet. This orthography is the same as that used for Standard Zhuang which is 1 of the 5 official languages of the People's Republic of China into which all national laws are translated and are included on paper bank notes. Rather than attempting to add kZhuang readings for over 10k encoded characters in one go, it is proposed that the readings be added in stages, starting with the 2,471 characters having GZ-sources and defined as follows:-

Property	kZhuang
Status	Provisional
Category	Readings
Delimiter	space
Syntax	[a-z]+*?
Description	The most customary Zhuang reading for this ideograph. Readings of words not part of the Standard Zhuang lexicon are suffixed by an asterisk. Among the sources used for the property data are the following: Ancient Zhuang Character Dictionary (古壮字字典), 1989, ISBN 7-5363-0614-8

In common with other Zhuang dictionaries the *Ancient Zhuang Character Dictionary* also notes if a word is not part of the Standard Zhuang lexicon, and it is proposed than kZhuang do likewise. In the *Ancient Zhuang Character Dictionary* readings of words not part of the Standard Zhuang lexicon have <方> written before them. Unihan should as indicate this in some way because the distinction is an integral part of Zhuang dictionaries. The use of an asterisk after the readings of words that are not a part of the standard Zhuang lexicon is a possible solution. A comparison of the different entries for two Zhuang characters meaning *sky* from Ext F, 𑜏 U+2D446 and 𑜏 U+2DC47 respectively, illustrates how this would work. The reading of the former is part of the Standard Zhuang lexicon but the latter is not (see figures 1 and 2).

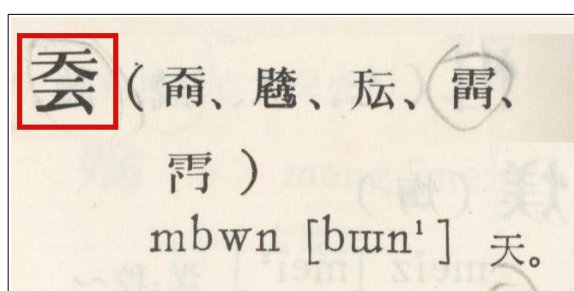


Figure 1: Ancient Zhuang Character Dictionary 'mbwn' page 322

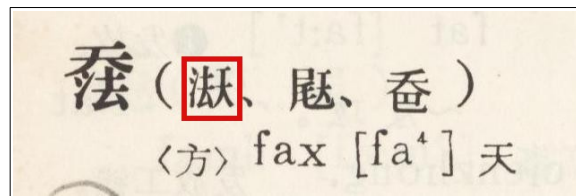


Figure 2: Ancient Zhuang Character Dictionary 'fax' page 156

𠵿 U+2D446 has source GZ-3222401, namely *Ancient Zhuang Character Dictionary* page 322 column 2 entry 4 character 1. The reading is Standard Zhuang so the Unihan kZhuang entry would be:-

U+2D446 kZhuang mbwn

𠵿 U+2DC47 has source GZ-1561402, namely *Ancient Zhuang Character Dictionary* page 156 column 1 entry 4 character 2. The reading is not Standard Zhuang so the Unihan kZhuang entry would be:-

U+2DC47 kZhuang fax*

This distinction is also noted in the standard reference *Sawloih Cuengh Gun (Zhuang-Chinese Dictionary)* which states that whilst both mean *sky* that *mbwn* is Standard Zhuang and *fax* is not (see figure 3).

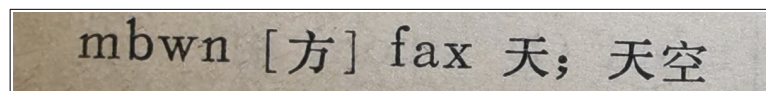


Figure 3: Sawloih Cuengh Gun(1984) 'mbwn' page 524

In the vast majority of cases the correct reading and whether or not it is part of the Standard Zhuang lexicon can easily be determined from the entry in *Ancient Zhuang Character Dictionary* but for 5 readings out of the 2,471 further investigation was required as shown below in Appendix 1.

For the initial list of 2,471 characters the value is based on the reading of GZ- source of the character, therefore each character has only one Zhuang reading even when the *Ancient Zhuang Character Dictionary* has several readings for the character. In the future some characters will require multiple Zhuang readings therefore the delimiter space is suggested.

Conclusion

The addition of kZhuang to Unihan at this time is logical and would help preserve useful information provided to IRG as a usual part of the encoding process. The 2,471 initial values proposed provide a good foundation on which to build. When more values are added the definition of kZhuang could be updated as required.

The author cordially invites those interested in helping in extending kZhuang to contact him via email knightleyjohnc@gmail.com or otherwise.

Acknowledgements:-

Thanks to Eiso Chan for discussions leading to this proposal, to Yi Bai for feedback on the data and to Ken Lunde for feedback on this document.

Appendix 1: Five readings that required further investigation

Readings 1-3:-

Whether or not a word is part of the Standard Zhuang lexicon is by no means naturally intuitive even for native speakers because the the Standard Zhuang lexicon includes words from different Zhuang languages. To determine whether or not a word is part of the Standard Zhuang lexicon is a matter of consulting appropriate reference works. Whilst the *Ancient Zhuang Character Dictionary* usually correctly indicates whether or not the readings are part of the Standard Zhuang lexicon, there are some exceptions. One such exception is *yoj*, to look, on page 510 (see figure 4).

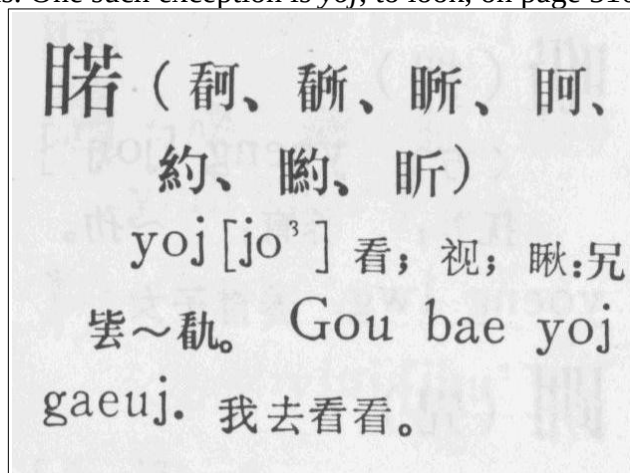


Figure 4: *Ancient Zhuang Character Dictionary* 'yoj' page 510

That the word *yoj*, to look, is not Standard Zhuang is explicitly stated twice in *Sawloih Cuengh Gun*. Firstly, under the standard *yawj*, where *yoj* is listed as the 5th of 8 non standard alternatives (see figure 5) and second, under *yoj*, where the reader is directed to the standard *yawj* (see figure 6)

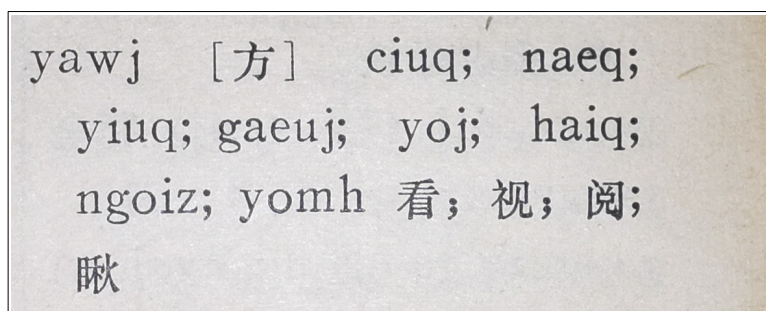


Figure 5: *Sawloih Cuengh Gun*(1984) 'yawj' page 734

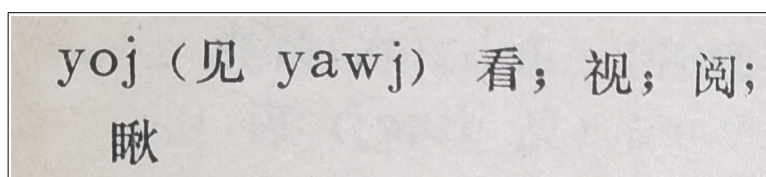


Figure 6: *Sawloih Cuengh Gun*(1984) 'yoj' page 745

This affects the values of 3 characters: GZ-5101302 𠵹 U+2DFCB, GZ-5101303 𠵺 U+2DFE1, and GZ-5101305 𠵻 U+2DFA8 should be

U+2DFCB	kZhuang	yoj*
U+2DFE1	kZhuang	yoj*
U+2DFA8	kZhuang	yoj*

and not

U+2DFCB	kZhuang	yoj
U+2DFE1	kZhuang	yoj
U+2DFA8	kZhuang	yoj

Readings 4 and 5:-

Due to slight printing errors, the correct reading of two other characters whilst not immediately obvious can be deduced from context:-

4. The correct reading for 鍬 U+30f41 is *gu* not *guh*. Though the header in *Ancient Zhuang Character Dictionary* shows *guh*, the ipa and the two examples show *gu* is in fact the correct reading (see figure 7) and a comparison to *Sawloih Cuengh Gun* also confirms that *gu* not *guh* is correct.

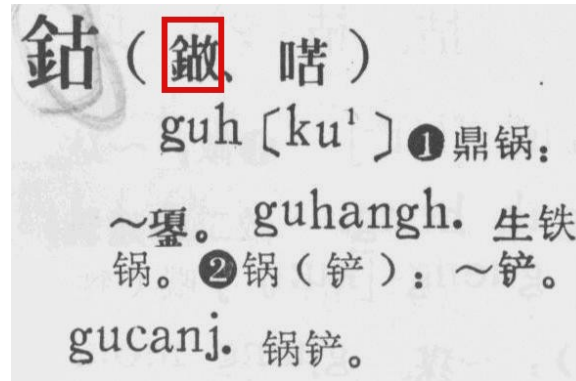


Figure 7: Ancient Zhuang Character Dictionary 'gu' page 205

5. Strictly speaking the reading for U+3193C of daem(lect) comes from not from GZ-0952501 itself but from the following entry which shows both the character and reading because GZ-0952501 itself only shows the character and has no reading. The evidence from the following entry was added via ORT (see <https://hc.jsecs.org/irg/ws2017/app/index.php?find=GZ-0952501>).