Proposed Changes to
the Unihan Database's `kMandarin` Field

John H. Jenkins

11 June 2022

Summary

I've done a comparison of Mandarin readings in the Unihan database with those in the 2003 edition of the *Shāngwù Xīn Zìdiǎn* (商務新字典), an intermediate, general-purpose dictionary published in Hong Kong. Where the two disagree, I have done a comparison with the values in the *Xiàndài Hànyǔ Cídiǎn* (现代汉语词典) and *Hànyǔ Dà Cídiǎn* (漢語大詞典). The former is the basis for the Unihan database's `kXHC1983` field, and the latter is a massive, comprehensive Chinese dictionary. We also checked the `kCantonese` field for cognates.

There are ten instances where, based on the readings in the *Shāngwù Xīn Zìdiǎn* and *Hànyǔ Dà Cídiǎn,* as well as the Cantonese cognates, it appears that the current `kMandarin` value does not reflect modern Chinese use.

Code Point	Character	kMandarin	Hanyu Da Cidian	Shangwu Xin Zidian	kCantonese
U+5A65	婥	nào	chuò	chuò	coek3
U+8F05	輅	hé	lù	lù	lou6
U+91D3	釓	qiú	gá	gá	gaa1
U+9208	鈈	pī	bù	bù	bat1
U+925A	鉚	liǔ	mǎo	mǎo	maau5
U+927B	鉻	luò	gé	gè	gok3
U+92E8	鋨	tiě	é	é	ngo4
U+9312	鋒	kē	ā	ā	aa1
U+937A	鍺	duǒ	[zhě]	zhě	ze2
U+499F	䦟	wěn	chuài	chuài	ceoi3

We recommend that in these ten instances, the kMandarin value be changed.

Notes:

None of these ten have entries in the *Xiàndài Hànyǔ Cídiǎn*.

In all cases, the current kCantonese value matches the Cantonese reading in the *Shāngwù Xīn Zìdiǎn*.

The *Hànyǔ Dà Cídiǎn* reading for U+937A 鍺 is in brackets because the primary meaning is given as "a variant of 堵." The first pronunciation given for U+937A 鍺 in its own right is "zhě."

It is rather late to be incorporating these changes in Unicode 15.0, given the need for additional vetting. The recommendation is that they be targeted to Unicode 16.0.