

Universal Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document
Title: On the Current Status of kCangjie Property
Source: Yang Jihai (楊霽海), Tan Keong Khon (陳強勤),
Woo Chi Chung (胡智聰), Jerry You (游程宇)
Status: Individual Contribution
Action: For consideration by UTC
Date: 2025-09-25

In reviewing the [kCangjie](#) field against the [“Cangjie 3rd-Generation Completion Project”](#) (倉頡三代補完計劃) and the [“Cangjie 5th-Generation Completion Project”](#) (倉頡五代補完計劃), we have identified several critical issues that make kCangjie outdated, inconsistent, and increasingly detached from present-day requirements.

1. Mixing of Generations

The mainstream Cangjie system has two major versions in active use: Cangjie 3rd-Generation and Cangjie 5th-Generation. However, kCangjie represents an unstructured mixture of both, without fully following either.

For example:

- Characters like 兔, 冤 are defined using 3rd-Generation rules (NUI, BNUI).
- Yet, characters such as 槐 instead follow 5th-Generation rules (DNAI).

Such cross-generation inconsistencies appear throughout the field.

2. Incorrect Encodings

Numerous characters are simply mis-encoded. Examples include: *(The left column under Proper Value is for Cangjie 3rd-Generation, and the right column is for Cangjie 5th-Generation; if there is only one column, it means the input code for two generations are the same.)*

UCS	Char.	Current Value	Proper Value	
U+3412	𠂇	SRN	SHRN	
U+3413	𠂈	MNRN	MRN	
U+365C	臺	YBG	YRBG	
U+3767	窰	JVOOG	JVOG	
U+4ACD	𨨗	LYMBC	LMYYC	LMSYC
U+3EB8	堡	OIMGI	ONMGI	
U+793B	𠂇	IF	INF	
U+94A1	钡	CBO	OPBO	

These errors illustrate the outdatedness of kCangjie and its divergence from modern input standards.

3. Unicode Mapping Errors

Some mappings between Cangjie and Unicode are historically flawed and remain unresolved.

For example:

- kCangjie defines 捏 (U+634F) as QHXM, but this code in fact corresponds to 捏 (U+63D1).
- The root cause: Cangjie 3rd-Gen manuals included 捏 but not 捏, while Big5 included 捏 but not 捏. As a result, 捏 was incorrectly mapped onto 捏, an error carried forward to the present day.

Such legacy mistakes demand urgent correction.

4. The “Duplicate Code” Problem

Originally, Cangjie input served not only as an IME but also as an internal coding system, which required strict uniqueness of codes. To handle duplicates, Chu Bong-Foo (朱邦復) introduced an X-prefix (重) system. For example:

- Both 𧈧 (U+8711) and 𧈩 (U+9CC9) had the code NMLMI, so 𧈩 was reassigned as XNMLM.

However, this approach is now problematic:

- (1) Users care about the original encoding, not artificial de-duplication.
- (2) Different code tables resolve duplicates differently, undermining consistency.
- (3) In the Unicode era, strict uniqueness is no longer necessary.
- (4) With the ongoing expansion of Unicode into Plane 2 and beyond, the X-prefix model is unsustainable.

We therefore recommend:

- Either abandoning the X-prefix for duplicates like 𧈩,
- Or at least retaining the original unmodified code alongside any de-duplication scheme.

5. Conclusion

The `kCangjie` field reflects an outdated, inconsistent, and error-prone system, rooted in legacy design constraints that are no longer relevant in today’s Unicode-based environment. Unless these structural flaws are addressed—mixing of generations, incorrect encodings, faulty mappings, and obsolete de-duplication—the field will remain unreliable and detached from the actual needs of modern users.

We hope to serve as volunteers to reorganize the current data of `kCangjie` property and add more data for more characters. If UTC agrees, we will submit the complete feedback and suggestions later. We suggest including Cangjie 3rd- and 5th-Generations input codes to `kCangjie` property at the same time with its source(s) following a colon like [kHangul](#) property.

We also have plans to review and check the Cangjie input codes used in HKSCS and MSCS. We will submit the feedback to UTC after finishing the review work.

(End of Document)

Appendix

There are five parts to introduce the background information of our organization works.

a. Brief History of the Cangjie Input Method

- 1977: Mr. Chu Bong-Foo introduced the first generation of the Cangjie input method.
- 1981: He released the second generation, which was used on Apple computers through an add-on called the 'Han Card' (漢卡).
- 1983: The third generation was launched and laid the foundation for the Chinese computing market.
- 1985: Chu introduced the fifth generation of Cangjie in the U.S. compared to the third generation. It modified some rules and added a small number of radicals.

For example, the character 面 (U+9762) in the third generation is decomposed as “丌, 口, 卜, 丨,” producing the code MWYL.

In the fifth generation, it's decomposed as “丌, 口, 𠂇, 丨,” producing the code MWSL.

Status:

Microsoft Windows includes “Microsoft Cangjie,” which is based on the third generation.

Apple iOS and macOS include Cangjie input methods compatible with both the third and fifth generations, meaning both MWYL and MWSL can be used to input the character 面.

The third and fifth generations are the most popular and useful for the end-users, this is the reason why we just focus on them.

b. ChineseCangjie (倉頡之友)

Background:

ChineseCangjie, originally named Cangjie Users Malaysia, was founded in 1997 by Mr. Tan Keong Khon (online name: Ahkhon). Under Chu Bong-Foo's commission, he promoted the Cangjie input method and compiled code tables.

Support from Chu Bong-Foo:

The “Chu Bong-Foo Studio” website repeatedly mentioned that the promotion of Cangjie was entrusted to ChineseCangjie, please see [here](#). Chu also provided financial support for server costs.

Status:

ChineseCangjie's fifth-generation code table has been compiled up to CJKUI Extension G and is included in popular input methods such as Google Gboard and Baidu IME.

c. Cangjie 3rd-Generation Completion Project (倉頡三代補完計劃)

This project is essentially a code table (a text file mapping Chinese characters to their Cangjie input codes for the third generation of the Cangjie input method).

Author: Yang Jihai (online names: 馬拉一個錘子 / ArthurMcArthur), forum administrator of ChineseCangjie.

Start: 2017.

Reason: In the past, expansions of the Cangjie 3rd-Generation code table were mostly community-driven, often inaccurate, and slow to update — frequently falling behind Unicode development.

Goal: Correct errors, improve glyph compatibility, and keep pace with Unicode.

Progress: The project now supports Unicode 17.0, including Extension J and compatibility characters. The basic block portion has been incorporated into Sogou Mobile Input.

d. Cangjie 5th-Generation Completion Project (倉頡五代補完計劃)

This project is essentially a code table (a text file mapping Chinese characters to their Cangjie input codes for the fifth generation of the Cangjie input method).

Author: Jackchows, a ChineseCangjie forum member.

Start: 2017.

Origin: Initially launched to extend the ChineseCangjie' fifth-generation table to cover Extensions E and F.

Goal: The same as the Cangjie 3rd-Generation Completion Project.

Progress: Currently supports Unicode 16.0, including CJKUI Extension I and compatibility characters.

Communication with Chu Bong-Foo Studio

2017: Jihai wrote to Chu Bong-fu studio for the first time to discuss the history of the Cangjie input method and issues (such as the right code of “𠂇”). I received a reply from Ms. Shen Hong-lian (沈紅蓮), who has been an important assistant to Mr. Chu during the creation of the Cangjie input method.

Later: I corresponded with the studio multiple times on different issues. They also kindly gifted us with a local edition of the Fifth Generation Cangjie Input Method Manual (published in 2006), which greatly aided my research.

e. Main Reference Materials of Cangjie 3 / 5 Completion Project:

- (1) 《零壹中文電腦叢書之八 倉頡第三代中文字母輸入法》 (1984, 全華)
- (2) 《零壹中文電腦叢書之九 增訂版倉頡第三代中文輸入法》 (1991, 全華)
- (3) 《第五代倉頡輸入法手冊》 (1993, 松崗; 1999, 文化傳信; 2006, 博碩)
- (4) 《零壹中文電腦叢書之七 標準倉頡第二代中文輸入法》 (1983, 全華)
- (5) 《內碼對照表》 (朱邦復工作室)
- (6) Correspondence with Ms. Shen Honglian, assistant to Chu Bong-Foo.