IP Feedback on the Chinese Root LGR (CLGR13)

**Last updated: May 22, 2019**

Contents

[1 Summary 3](#_Toc9447520)

[2 Definitions 4](#_Toc9447521)

[2.1 CLGR13 4](#_Toc9447522)

[2.2 IICORE collection 5](#_Toc9447523)

[2.3 MSR-4 CJK repertoire 5](#_Toc9447524)

[2.4 dotAsia LGR 5](#_Toc9447525)

[2.5 CDNC-2005 5](#_Toc9447526)

[2.6 CDNC-2018 5](#_Toc9447527)

[2.7 Unihan 5](#_Toc9447528)

[3 Repertoire considerations 6](#_Toc9447529)

[4 Variant set differences by type of difference 6](#_Toc9447530)

[4.1 Additional repertoire (CLGR13 adds 2 code points not in dotAsia) 6](#_Toc9447531)

[4.2 Code points part of CDNC 2018 and dotAsia but treated differently 6](#_Toc9447532)

[4.2.1. Introduction 6](#_Toc9447533)

[4.2.2. Summary of the list of additions treated differently – 35 items 7](#_Toc9447534)

[4.2.3. Detailed list of differences for CDNC2018/dotAsia additions still under review – 3 sets 9](#_Toc9447535)

[4.3 Code points new to CLGR13, not in CDNC2018, also part of dotAsia but treated differently 14](#_Toc9447536)

[4.3.1. Introduction 14](#_Toc9447537)

[4.3.2. Summary of the list of additions treated differently – 19 items 14](#_Toc9447538)

[4.3.3. Detailed list of remaining differences for dotAsia additions – 4 sets 15](#_Toc9447539)

[4.4 Variant sets corresponding to visual exchangeability 19](#_Toc9447540)

[5 Issues with XML files 22](#_Toc9447541)

[6 Issues with the Proposal document 23](#_Toc9447542)

[6.1 General 23](#_Toc9447543)

[6.2 Specific feedback concerning CLGR13 section about pre-integration 23](#_Toc9447544)

[6.3 Mechanism for reducing multiple allocatable labels 25](#_Toc9447545)

[6.4 Visually Identical Characters 26](#_Toc9447546)

[7 Notation and explanation used in the description of the variant set differences 28](#_Toc9447547)

[7.1 Table format for variant sets 28](#_Toc9447548)

[7.2 Source information 28](#_Toc9447549)

[8 Detailed review of variant set differences 29](#_Toc9447550)

[8.1 Review of variant sets for additional HKSCS characters beyond dotAsia – 2 sets 29](#_Toc9447551)

[8.2 Detailed list of differences for CDNC2018/dotAsia additions – 35 sets 34](#_Toc9447552)

[8.3 Detailed list of differences for dotAsia additions – 19 sets 54](#_Toc9447553)

[9 Background concerning visual exchangeability 70](#_Toc9447554)

[10 References 72](#_Toc9447555)

PART I — Overview and summary of IP recommendations

# Summary

This document presents Integration Panel recommendations for the Chinese Root Zone LGR proposal together with an analysis of the content of the current Chinese LGR as specified by the Chinese Generation Panel in terms of repertoire and variant sets. The latest draft of that Chinese LGR (CLGR13) 0Fis represented by the following files:

* CGP-LGR-Proposal-1.4-201904, referred to as [Proposal] below
* Appendix A CGP Repertoire [201902].xlsx,
* Appendix B JGP Repertoire [201703].xlsx,
* Appendix C KGP Hanja Repertoire [201703].xlsx,
* Appendix D CGP Variant Mappings [201904].xlsx,
* Appendix E CGP Internal Review [201904].xlsx,
* Appendix F CGP and KGP Pre-Integration on 258 Variant Groups [201902].xlsx,
* Appendix G Appendix G CDNC IDN Table 2018 - Unicode.docx,
* Appendix G Appendix G CDNC IDN Table 2018 - Unicode.txt,
* Appendix H KGP Hanja Variant Mappings [201703].xlsx,
* Appendix I CGP Variant Mappings Matching Existing Practice [201904].xlsx,
* Appendix J CGP Variant Mappings Differ from Existing Practice [201902].xlsx,
* Appendix K CGP Multiple Mappings [201904].xlsx,
* Appendix L CDNC IDN Table.xml
* Appendix M dotAsia IDN Table.xml,
* Appendix N CGP Internal Review.xml,
* Appendix O C&K Pre-Integration.xml,
* Appendix P KGP Variant Groups.xml
* Appendix Q 260 unacceptable variant groups from KGP [201603].xlsx
* Appendix R 150 group splitting proposal by KGP [201609].xlsx
* Appendix S Statistic of 258 unacceptable variant groups in CN-TW-网址 reg database [201609].xlsx
* Appendix T Final consensus doc on 146 unacceptable groups [201703].docx
* Proposed-LGR-Hani-1.4-201904.xml,
* Labels-Hani-20180824.txt

(additional files were provided but are not discussed in this document)

The following list summarizes the analysis:

**Repertoire**

* The repertoire contains 19,685 code points, all CJK ideographs, very close to the size of MSR-3 Hanzi set (19,853). It is a full superset of the dotAsia CJK ideograph repertoire that includes 19,683 code points.
* The repertoire includes 2 HKSCS (Hong Kong Supplementary Character Set) characters not part of the dotAsia; these were recent additions to MSR (new to MSR-3) and have been added to CDNC-2018.
* CDNC-2018 itself contains 19,563 CJK ideographs, only 112 less than CLGR13.

**Variant sets**

* The differences in variant sets are discussed in section 4.1 (related to the two new repertoire additions not in dotAsia), 4.2 (related to recent additions common to dotAsia and CDNC-2018), 4.3 (related to additions included in dotAsia but not CDNC-2018), and 4.4 (related to visual exchangeability)).
* Sections 4.1, 4.2, and 4.3 have been extensively reviewed by our IP advisor on CJK issues, Dr. Lu Qin, Hong Kong Polytechnic University, who is a leading expert on CJK encoding matters.
* To facilitate comparison of various variant sets, all mapping differences resulting from mitigation used to decrease multiple allocatable variants may have been removed from some LGRs.

**Reduction of multiple allocatable variants**

IP has reviewed again the mechanism developed by CGP to reduce these multiple variants. This has resulted in a number of recommendations detailed in section 6.3.

**Integration Panel recommendations:**

* No further comments on repertoire of 19,685 CJK Hanzi characters, although CGP may consider suggesting that CDNC-2018 be aligned with CLGR13 in term of repertoire and variant sets.【forwarded to CDNC mailing list now and waiting for their collective decision】
* Replace reference to MSR-3 by MSR-4. 【done】
* Review variant set differences detailed in sections 4.2, and 4.3 and make changes as appropriate. Note that for most of those differences (especially section 4.2), the IP is suggesting using the CLGR13 definitions instead of dotAsia.
* IP recommend adding the 3 variant sets as defined in section 4.4 concerning visually exchangeable characters.
* Follow IP suggestions concerning multiple allocatable variants as appropriate.
* If electing to not follow a recommendation, please provide a documented rationale for keeping the original.
* Follow the recommendations made in section 5 and 6 concerning the CLGR13 XML file and the [Proposal] document.

For details of and rationale for these recommendations, see the IP recommendations and comments in the remainder of the document.

Sections 3 and 4 address the technical content of the proposal (repertoire and variant sets), while section 5 covers the XML format, and section 6 covers the [Proposal] document itself. Section 7 describes the syntax used to review the variant sets, section 8 contains the full set of variant set differences between CLGR13 and dotAsia and is for reference use. Section 9 contains more background information about visual variants. Section 10 contains various references used in these considerations.

# Definitions

## CLGR13

The term (CLGR13) represents the Proposed Chinese root LGR under review here, both in terms of repertoire and variant sets defined in the XML file. The term CLGR12 may be used to represent an earlier version of the Chinese LGR.

## IICORE collection

The International Ideographs Core (IICORE) is a fixed collection of CJK Ideographic code points deemed essential to all IRG Asian constituencies except Vietnam (a total of 7 sources). It contains 9,810 code points and is part of both ISO/IEC 10646 and Unicode. It was created by IRG based on priority (A to C, A being the highest) among its 7 sources.

## MSR-4 CJK repertoire

The CJK repertoire in MSR-4 consists of 19,853 CJK Unified Ideographs, corresponding to the union of the following sub-repertoires:

1. dotAsia Japanese <https://www.iana.org/domains/idn-tables/tables/asia_ja_1.1.txt>
2. dotAsia Chinese <https://www.iana.org/domains/idn-tables/tables/asia_zh_1.1.txt>
3. IICORE as defined in Unicode 6.3
4. Code points U+3A5C, U+58B5, U+9DC0, and U+20B9F.

The dotAsia Chinese repertoire is itself a union of repertoires from various Chinese sources such as China PRC, Hong Kong SARs, and Taiwan.

Note that MSR-4 also contains a few code points that have the ‘Han’ extended script property but are not considered CJK Ideographs (for example U+3005 IDEOGRAPHIC ITERATION MARK and U+3006 IDEOGRAPHIC CLOSING MARK).

## dotAsia LGR

The dotAsia (ZH) IDN table is available at <https://www.iana.org/domains/idn-tables/tables/asia_zh_1.1.txt>. An XML-format of that file is provided as Appendix M to CLGR13.

The dotAsia repertoire is fully included in CLGR13. CLGR13 adds just 2 code points: U+3A5C and U+58B5.

## CDNC-2005

This is the IDN table that can be constructed by merging the IDN tables created for .cn and .tw in 2005 and still in use today. That table contains 19,520 CJK Hanzi characters. There is no xml description of that table. 【has provided as Appendix ?? to CLGR14】

## CDNC-2018

This IDN table results from further refinement to the original CDNC-2005 table. While an online reference exists at <http://www.cdnc.asia/file/unicode-1-2.txt> , there is still no strong evidence of use. That table has a XML representation provided in Annex L of the [Proposal]. The table contains 19,563 CJK Hanzi characters. The repertoire is a clean superset of CDNC-2005, but there are incompatible differences in the variant sets of both tables. 【Just restated the concerns from CGP and IP, urging CDNC to take action ASAP】

## Unihan

The Unihan database at <http://www.unicode.org/charts/unihan.html> is a Unicode Standard component containing information related to all CJK Ideographs. That information includes sources, variants, dictionaries, etc. As such, it is a very useful tool for validating the CLGR13 content.

Part II — IP Recommendations in Detail

# Repertoire considerations

The IP has no further comment on the CLGR repertoire and consider the repertoire as presently defined acceptable.

# Variant set differences by type of difference

## Additional repertoire (CLGR13 adds 2 code points not in dotAsia)

These two cases represent variant sets modified by the addition of two new characters to CLGR on top of dotAsia. The IP has made recommendation for both variant sets and these recommendations have been endorsed by its own CJK expert. Note that these 2 HKIRC characters are also part of CDNC 2018. Both cases involve multiple causes for the differences (addition of these 2 characters, additions after CDNC 2018, etc.) which are shown in the table below (HKIRC characters shown in red in the table).

| **No** | **CLGR13 Variant set glyphs and code points** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 㧛 39DB | 㩜 3A5C | 揽63FD | 擥64E5 | 攬652C |  |  |  |
| 2 | 坛575B | 埮57EE | 墰58B0 | 墵58B5 | 壇58C7 | 壜58DC | 罈7F48 | 罎7F4E |

**IP Recommendation**

* After review, IP is endorsing the CLGR 13 version.

Section 8.1 gives a detailed review of these two sets and rationale for the IP recommendation. 【thanks】

## Code points part of CDNC 2018 and dotAsia but treated differently

### Introduction

The variant sets in this section have code points that are common to the CDNC2018 and dotAsia repertoires, but the chosen variant mappings differ from those used in the original dotAsia set. To show the differences, variant sets may be listed twice, once for CLGR13 and once for dotAsia. See section 7 for the format of the notation and explanation used in the descriptions used in this section.

Most of these cases correspond to the 43 new Chinese characters added recently in the proposed CDNC 2018 as requested by HKIRC (see section 5.2.1 of the CLGR13 Proposal). However, 4 cases with no new code points corresponding to the existing CDNC practice (CDNC 2005) are also included in this list of variant sets. There is one hybrid case where the variant set also contains one code point: U+58B5 not included in CDNC 2018 and is covered in section 4.1.

These cases are all described in a different format in CLGR13 Annex D worksheet 6.2.1.1-75, which contains 75 lines of differences in variant sets. The format used in that worksheet is a format similar to RFC 3743, which does not directly translate to variant sets, thus the discrepancy between 75 entries in that file and the number of 35 variant sets below. In addition, these 75 lines do not capture all lines of a given variant set (lack of symmetry). Finally, because some of these 75 entries are interacting with additions beyond CDNC2018, they are described in the following section (4.3).

When refining the analysis with the help of a CJK expert, the IP has found that many of the differences are the result of conflicting opinion on the relative weight of a new HKSCS character (typically ‘traditional’ or ‘both’) with respect to an existing character which is itself a traditional variant (or a singleton reflexive). In general, dotAsia treats the new character as equal while CLGR13 tends to treat the existing character as ‘preferred’. After further review by our CJK expert, the CLGR13 version was in general found to be preferable and therefore accepted.

In most cases, IP has provided a ‘proposed’ disposition to either accept CLGR 13 or preserve dotAsia definition. In some rare cases, another alternative is proposed.

**IP Recommendation**

* Review among the 35 items summarized in the next section, the 4 cases where IP still disagree with the CGP proposal and provide additional justification for the change.

Section 8.2 gives a detailed review of all 35 sets and rationale for the IP recommendations.

### Summary of the list of additions treated differently – 35 items

Legend:

* Code points in red represent common CDNC2018/dotAsia additions,
* Code point in blue represents CDNC2018 addition not in dotAsia,
* Green-highlighted code points represents either singleton or additional variant sets,
* The last column (Note) describes the current disposition of the differences (either to adopt the dotAsia variant sets or the CLGR13 variant sets). If a ‘\* is presents, it represents a third choice.
* The details are shown in the next section using the same index number.
* **Bold** lines show lines where the CLGR13 set is not recommended and are detailed using the numbering shown in parenthesis (1 to 3) in the next section.

| **No** | **Variant set(s) glyphs and code points** | | | | | | | | **Note** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 假5047 | 叚53DA |  |  |  |  |  |  | CLGR13 |
| 2 | 刊520A | 刋520B | 栞681E |  |  |  |  |  | CLGR13 |
| 3 | 劤52A4 | 劲52B2 | 勁52C1 |  |  |  |  |  | CLGR13 |
| 4 | 发53D1 | 彂5F42 | 発767A | 發767C | 髪9AEA | 髮9AEE |  |  | CLGR13 |
| 5 | 吻543B | 呅5445 | 呡5461 | 脗8117 |  |  |  |  | CLGR13 |
| 6 | 园56ED | 園5712 | 薗8597 |  |  |  |  |  | CLGR13 |
| 7 | 坛575B | 埮57EE | 墰58B0 | 墵58B5 | 壇58C7 | 壜58DC | 罈7F48 | 罎7F4E | CLGR13 |
| 8 | 埗57D7 | 埠57E0 |  |  |  |  |  |  | CLGR13 |
| **9** | 嬎**5B0E** | 嬔**5B14** |  |  |  |  |  |  | **\* (1)** |
| **10** | 宁**5B81** | 寍**5BCD** | 寕**5BD5** | 寗**5BD7** | 寜**5BDC** | 寧**5BE7** | 甯**752F** |  | **\* (2)** |
| 11 | 宝5B9D | 寳5BF3 | 寶5BF6 | 珤73E4 |  |  |  |  | CLGR13 |
| 12 | 帚5E1A | 箒7B92 | 菷83F7 |  |  |  |  |  | CLGR13 |
| 13 | 慼617C | 慽617D | 戚621A |  |  |  |  |  | CLGR13 |
| 14 | 抬62AC | 擡64E1 |  |  |  |  |  |  | CLGR13 |
| 15 | 枣67A3 | 栆6806 | 棗68D7 |  |  |  |  |  | CLGR13 |
| 16 | 柄67C4 | 棅68C5 |  |  |  |  |  |  | CLGR13 |
| 17 | 椀6900 | 盌76CC | 碗7897 | 鋺92FA |  |  |  |  | CLGR13 |
| 18 | 橓6A53 | 蕣8563 |  |  |  |  |  |  | CLGR13 |
| 19 | 泷6CF7 | 滝6EDD | 瀧7027 |  |  |  |  |  | CLGR13 |
| 20 | 爀7200 | 赫8D6B |  |  |  |  |  |  | CLGR13 |
| 21 | 獏734F | 貘8C98 |  |  |  |  |  |  | CLGR13 |
| 22 | 珡73E1 | 琴7434 |  |  |  |  |  |  | CLGR13 |
| 23 | 痫75EB | 癇7647 | 癎764E |  |  |  |  |  | CLGR13 |
| 24 | 睹7779 | 覩89A9 |  |  |  |  |  |  | CLGR13 |
| 25 | 秕79D5 | 粃7C83 |  |  |  |  |  |  | CLGR13 |
| 26 | 筯7B6F | 箸7BB8 |  |  |  |  |  |  | CLGR13 |
| **27** | 範**7BC4** | 范 **8303** |  |  |  |  |  |  | **\* (3)** |
| 28 | 綿7DBF | 緜7DDC | 绵7EF5 |  |  |  |  |  | dotAsia |
| 29 | 膳81B3 | 饍994D |  |  |  |  |  |  | dotAsia |
| 30 | 蓚84DA | 蓨84E8 |  |  |  |  |  |  | dotAsia |
| 31 | 転8EE2 | 轉8F49 | 转8F6C |  |  |  |  |  | CLGR13 |
| 32 | 边8FB9 | 辺8FBA | 邉9089 | 邊908A |  |  |  |  | CLGR13 |
| 33 | 駅99C5 | 驛9A5B | 驿9A7F |  |  |  |  |  | CLGR13 |
| 34 | 鴟9D1F | 鵄9D44 | 鸱9E31 |  |  |  |  |  | CLGR13 |
| 35 | 齡9F61 | 齢9F62 | 龄9F84 |  |  |  |  |  | CLGR13 |

Note that the Annex D worksheet 6.2.1.1-75 contains some possible errors: 【accepted by CGP】

* The line entries corresponding to U+51C0, U+51C8, U+6D44, U+6DE8 correspond to a variant set unchanged by these HKIRC additions. It is further modified in CLGR13 by the code point U+701E, but with no impact on mapping between pre-existing code points. That modified variant set is described in section 4.3.
* The line entries corresponding to U+840C and U+8544 correspond to a variant set unchanged by these HKIRC additions. It is further modified in CLGR13 by the code point U+8420 but with no impact on mapping between pre-existing code points. That modified variant set is described in section 4.3.
* The following line:

12616 82CF 苏 … 苏(82CF) 苏(82CF) 蘇(8607) 甦(7526) 囌(56CC)甦(7526)苏(82CF)蘇(8607)蘓(8613)

concerns a variant set {U+56CC, U+7526, U+82CF, U+8607, U+8613} fully included in CDNC2018 and identical to the dotAsia version, therefore it is unclear why it was included in that worksheet.

All the summary items are captured by Annex D worksheet 6.2.1.1-75 with the exception of item #27 (範7BC4, 范 8303) which is likely a typo.

### Detailed list of differences for CDNC2018/dotAsia additions still under review – 3 sets

The following numbered items list all the information available to the IP at this point for each of the differences still under review and its recommendations based on evidence and the opinion of the IP CJK expert.

The code points U+5B0E and U+5B14 form their own variant set in dotAsia. In CLGR13, these two code points are singletons with reflexive variant of type ‘r-both’.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5B0E | 嬎 | 5B0E | 嬎 | ≡ | r-both |  | identity |
| 5B0E | 嬎 | 5B14 | 嬔 | ↔ | blocked |  |  |
| 5B14 | 嬔 | 5B14 | 嬔 | ≡ | r-both |  | identity |

Unihan does not provide any clues concerning this (no variant set for any of these code points). The glyph differences appear minuscule which may have been a factor in the treatment as blocked variant by dotAsia.

Although CLGR13 is identical to CDNC-2018, this case is also a case of visual variant (see section 4.5).

Our CJK expert and IP are in favor of keeping the variant set but with some way to have a way to have an allocatable label in at least one direction (making the pair allocatable in both directions is difficult to do within the CJK variant set mapping system). Given that U+5B14 is used more than U+5B0E a possible solution is:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5B0E | 嬎 | 5B0E | 嬎 | ≡ | r-neither |  | identity |
| 5B0E | 嬎 | 5B14 | 嬔 | **→** | both |  |  |
| **←** | blocked |
| 5B14 | 嬔 | 5B14 | 嬔 | ≡ | r-both |  | identity |

This allows U+5B14 to be an allocatable variant of U+5B0E, while U+5B0E is a blocked variant of U+5B14.

**Conclusion:** IP is requesting CGP to review the IP suggestion. 【To be discussed 】

The code point U+5BD7 was included in CLGR13 because of its membership in IICORE HKSCS but is treated differently between CLGR13 (1st) and dotAsia (2nd). In DotAsia, U+752F is a singleton reflexive variant of type ‘r-both’. In CLGR13, U+5BD7 has different variant mappings, and U+752F is part of this variant set.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5B81 | 宁 | 5B81 | 宁 | ≡ | r-both |  | identity |
| 5B81 | 宁 | 5BCD | 寍 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5B81 | 宁 | 5BD5 | 寕 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5B81 | 宁 | 5BD7 | 寗 | ↔ | blocked |  |  |
| 5B81 | 宁 | 5BDC | 寜 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5B81 | 宁 | 5BE7 | 寧 | **→** | trad |  |  |
| **←** | simp |  |  |
| 5B81 | 宁 | 752F | 甯 | ↔ | blocked |  |  |
| 5BCD | 寍 | 5BCD | 寍 | ≡ | r-trad |  | identity |
| 5BCD | 寍 | 5BD5 | 寕 | ↔ | blocked |  |  |
| 5BCD | 寍 | 5BD7 | 寗 | ↔ | blocked |  |  |
| 5BCD | 寍 | 5BDC | 寜 | ↔ | blocked |  |  |
| 5BCD | 寍 | 5BE7 | 寧 | ↔ | blocked |  |  |
| 5BCD | 寍 | 752F | 甯 | ↔ | blocked |  |  |
| 5BD5 | 寕 | 5BD5 | 寕 | ≡ | r-neither |  | identity |
| 5BD5 | 寕 | 5BD7 | 寗 | ↔ | blocked |  |  |
| 5BD5 | 寕 | 5BDC | 寜 | ↔ | blocked |  |  |
| 5BD5 | 寕 | 5BE7 | 寧 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 5BD5 | 寕 | 752F | 甯 | ↔ | blocked |  |  |
| 5BD7 | 寗 | 5BD7 | 寗 | ≡ | r-both |  | identity |
| 5BD7 | 寗 | 5BDC | 寜 | ↔ | blocked |  |  |
| 5BD7 | 寗 | 5BE7 | 寧 | ↔ | blocked |  |  |
| 5BD7 | 寗 | 752F | 甯 | ↔ | blocked |  |  |
| 5BDC | 寜 | 5BDC | 寜 | ≡ | r-neither |  | identity |
| 5BDC | 寜 | 5BE7 | 寧 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 5BDC | 寜 | 752F | 甯 | ↔ | blocked |  |  |
| 5BE7 | 寧 | 5BE7 | 寧 | ≡ | r-trad |  | identity |
| 5BE7 | 寧 | 752F | 甯 | ↔ | blocked |  |  |
| 752F | 甯 | 752F | 甯 | ↔ | r-both |  | identity |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5B81 | 宁 | 5B81 | 宁 | ≡ | r-both |  | identity |
| 5B81 | 宁 | 5BCD | 寍 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5B81 | 宁 | 5BD5 | 寕 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5B81 | 宁 | 5BD7 | 寗 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5B81 | 宁 | 5BDC | 寜 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5B81 | 宁 | 5BE7 | 寧 | **→** | trad |  |  |
| **←** | simp |  |  |
| 5BCD | 寍 | 5BCD | 寍 | ≡ | r-trad |  | identity |
| 5BCD | 寍 | 5BD5 | 寕 | ↔ | blocked |  |  |
| 5BCD | 寍 | 5BD7 | 寗 | ↔ | blocked |  |  |
| 5BCD | 寍 | 5BDC | 寜 | ↔ | blocked |  |  |
| 5BCD | 寍 | 5BE7 | 寧 | ↔ | blocked |  |  |
| 5BD5 | 寕 | 5BD5 | 寕 | ≡ | r-neither |  | identity |
| 5BD5 | 寕 | 5BD7 | 寗 | ↔ | blocked |  |  |
| 5BD5 | 寕 | 5BDC | 寜 | ↔ | blocked |  |  |
| 5BD5 | 寕 | 5BE7 | 寧 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 5BD7 | 寗 | 5BD7 | 寗 | ≡ | r-neither |  | identity |
| 5BD7 | 寗 | 5BDC | 寜 | ↔ | blocked |  |  |
| 5BD7 | 寗 | 5BE7 | 寧 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 5BDC | 寜 | 5BDC | 寜 | ≡ | r-neither |  | identity |
| 5BDC | 寜 | 5BE7 | 寧 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 5BE7 | 寧 | 5BE7 | 寧 | ≡ | r-trad |  | identity |

Unihan has a variant relationship for the set: U+5B81, U+5BCD, U+5BD5, U+5BDC, and U+5BE7 and another variant relationship for the set U+5BD7, U+752F.

Based on this, it would seem preferable to remove U+5BD7 from the previous set and make an additional variant set only including U+5BD7 and U+752F. The two sets look like:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5B81 | 宁 | 5B81 | 宁 | ≡ | r-both |  | identity |
| 5B81 | 宁 | 5BCD | 寍 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5B81 | 宁 | 5BD5 | 寕 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5B81 | 宁 | 5BDC | 寜 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5B81 | 宁 | 5BE7 | 寧 | **→** | trad |  |  |
| **←** | simp |  |  |
| 5BCD | 寍 | 5BCD | 寍 | ≡ | r-trad |  | identity |
| 5BCD | 寍 | 5BD5 | 寕 | ↔ | blocked |  |  |
| 5BCD | 寍 | 5BDC | 寜 | ↔ | blocked |  |  |
| 5BCD | 寍 | 5BE7 | 寧 | ↔ | blocked |  |  |
| 5BD5 | 寕 | 5BD5 | 寕 | ≡ | r-neither |  | identity |
| 5BD5 | 寕 | 5BDC | 寜 | ↔ | blocked |  |  |
| 5BD5 | 寕 | 5BE7 | 寧 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 5BDC | 寜 | 5BDC | 寜 | ≡ | r-neither |  | identity |
| 5BDC | 寜 | 5BE7 | 寧 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 5BE7 | 寧 | 5BE7 | 寧 | ≡ | r-trad |  | identity |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5BD7 | 寗 | 5BD7 | 寗 | ≡ | r-both |  | identity |
| 5BD7 | 寗 | 752F | 甯 | ↔ | blocked |  |  |
| 752F | 甯 | 752F | 甯 | ↔ | r-both |  | identity |

**Conclusion:** IP is requesting CGP to review the IP suggestion.【To be discussed 】

The variant set {U+7BC4, U+8303} is treated differently between CLGR13 (1st table) and both dotAsia and CDNC2018 (2nd table).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7BC4 | 範 | 7BC4 | 範 | ≡ | r-trad |  | identity |
| 7BC4 | 範 | 8303 | 范 | **→** | simp |  |  |
| **←** | trad-1 |  |  |
| 8303 | 范 | 8303 | 范 | ≡ | trad-2 |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7BC4 | 範 | 7BC4 | 範 | ≡ | r-trad |  | identity |
| 7BC4 | 範 | 8303 | 范 | **→** | simp |  |  |
| **←** | trad |  |  |
| 8303 | 范 | 8303 | 范 | ≡ | r-both |  | identity |

The CLGR 13 table has some format issue (‘trad-2’ is not a reflexive type), and ‘r-both-mt’ may have been the intended mapping (this was the case in CLGR12). However, the use of ‘trad-1’ and ‘trad-2’ may have been a trial at having two ‘trad’ mappings.

**Conclusion:** This needs further review. 【Accepted】

## Code points new to CLGR13, not in CDNC2018, also part of dotAsia but treated differently

### Introduction

The variant sets in this section have code points that are common to both the CLGR13 and dotAsia repertoires but are not in CDNC2018, for which the chosen variant mappings differ from those used in the original dotAsia set. To show the differences, variant sets may be listed twice, once for CLGR13 and once for dotAsia. See section 7 for the format of the notation and explanation used in the descriptions used in this section.

All these cases correspond to the BMP characters added to CLGR13 that are not in CDNC 2018 (53 characters described in section 5.2.2.1 of the CLGR13 proposal and 7 characters described in section 5.2.2.2 of the same proposal). The variant sets corresponding to the 62 characters located in the Supplement Ideographic Plane (SIP) are using the dotAsia version without modification.

These cases are all described in a different format in CLGR13 Annex D worksheet 6.2.1.3-34 (it should be 6.2.1.3-44) which contains 44 lines of differences in variant sets. The format used in that worksheet is a format similar to RFC 3743 which does not simply translate to variant set numbers, thus the discrepancy between 34 and the number 19 below. In addition, these 44 lines do not capture all lines of a given variant set.

As in previous section, the IP has found that many of the differences are the result of conflicting opinion on the relative weight of a new HKSCS characters.

**IP recommendation:**

* Please review these variant set differences and resolve them based on consensus among experts from different regions.
* Whether making changes or keeping the original mappings, please document a rationale.

Section 8.3 gives a detailed review of all 19 sets and rationale for the IP recommendations.

### Summary of the list of additions treated differently – 19 items

Legend:

* Code points in red represent dotAsia (but not in CDNC 2018) additions,
* Code point in blue represents CDNC2018/dotAsia addition (hybrid cases),
* In Item 1, U+3A5C is part of CDNC 2018 but not part of dotAsia; that variant set is described in 4.1,
* Green-highlighted code points represents either singleton or additional variant sets,
* The last column (Note) describes the current disposition of the differences (either to adopt the dotAsia variant sets or the CLGR13 variant sets). If a ‘\*’ is presents, it represents a third choice or undetermined.
* The details are shown in the next section using the same index number.
* **Bold** lines show lines where the CLGR13 set is not recommended and are detailed using the numbering shown in parenthesis (1 to 4) in the next section.

| **No** | **Variant set(s) glyphs and code points** | | | | | | | | | **Note** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 㧛39DB | 㩜3A5C | 揽63FD | 擥64E5 | 攬652C |  |  |  |  | CLGR13 |
| **2** | 㨘**3A18** | 擤**64E4** |  |  |  |  |  |  |  | **dotAsia (1)** |
| 3 | 㩒3A52 | 擒64D2 |  |  |  |  |  |  |  | CLGR13 |
| **4** | 䱽**4C7D** | 䲝**4C9D** | 鯧**9BE7** | 鲳**9CB3** |  |  |  |  |  | **\* (2)** |
| **5** | 净**51C0** | 凈**51C8** | 浄**6D44** | 淨**6DE8** | 瀞**701E** |  |  |  |  | **dotAsia (3)** |
| 6 | 弄5F04 | 挵6335 | 衖8856 |  |  |  |  |  |  | CLGR13 |
| 7 | 总603B | 惣60E3 | 揔63D4 | 捴6374 | 搃6403 | 摠6460 | 総7DCF | 縂7E02 | 總7E3D | CLGR13 |
| 8 | 扬626C | 揚63DA | 敭656D | 颺98BA | 飏98CF |  |  |  |  | CLGR13 |
| 9 | 捿637F | 栖6816 | 棲68F2 |  |  |  |  |  |  | CLGR13 |
| 10 | 晧6667 | 暠66A0 | 皓7693 | 皜769C |  |  |  |  |  | CLGR13 |
| 11 | 灶7076 | 竃7AC3 | 竈7AC8 |  |  |  |  |  |  | CLGR13 |
| 12 | 炼70BC | 煉7149 | 錬932C | 鍊934A |  |  |  |  |  | CLGR13 |
| 13 | 琉7409 | 瑠 7460 |  |  |  |  |  |  |  | CLGR13 |
| 14 | 琍740D | 璃7483 | 瓈74C8 |  |  |  |  |  |  | CLGR13 |
| 15 | 畺757A | 疆7586 |  |  |  |  |  |  |  | CLGR13 |
| 16 | 肕8095 | 靭976D | 靱9771 | 韌97CC | 韧97E7 |  |  |  |  | CLGR13 |
| **17** | 萌**840C** | 萠**8420** | 蕄**8544** |  |  |  |  |  |  | **dotAsia (4)** |
| 18 | 鉄9244 | 銕9295 | 鐵9435 | 铁94C1 |  |  |  |  |  | CLGR13 |
| 19 | 飆98C6 | 飇98C7 | 飈98C8 | 飙98D9 | 飚98DA |  |  |  |  | CLGR13 |

All the items in the summary are captured in Annex D worksheet 6.2.1.1-75 and 6.2.1.3-34. Note that the worksheet 6.2.1.3-34 has the following error:

* Three entries for the code points U+55B0, U+98E1, and U+9910 correspond to a variant set which is now identical between dotAsia and CLGR12, therefore they should be deleted.【accepted by CGP 】

### Detailed list of remaining differences for dotAsia additions – 4 sets

The following numbered items list all the information available to the IP at this point for each of the remaining differences and its recommendations based on evidence and the opinion of the IP CJK expert.

The code point U+3A18 was included in CLGR13 because of its membership in IICORE HKSCS but is has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 3A18 | 㨘 | 3A18 | 㨘 | ≡ | r-trad |  | identity |
| 3A18 | 㨘 | 64E4 | 擤 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 64E4 | 擤 | 64E4 | 擤 | ≡ | r-both |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 3A18 | 㨘 | 3A18 | 㨘 | ≡ | r-both |  | identity |
| 3A18 | 㨘 | 64E4 | 擤 | ↔ | blocked |  |  |
| 64E4 | 擤 | 64E4 | 擤 | ≡ | r-both |  | identity |

Unihan in the entry for U+3A18 mentions in its kDefinition field that it is “(same as U+64E4 擤) to blow the nose with the fingers; (Cant.) to scour; to rebuke; to hit with a ball”; but it has no traditional/simplified mapping.

In the 2nd table above, dotAsia treats them as symmetric. In the 1st table above, CLGR13, by making U+3A18 a reflexive ‘r-trad’’ and making U+3A18**→**U+64E4 ‘simp’ (which is odd but consistent with ‘r-trad’) is favoring U+64E4 by allowing U+64E4 to be a ‘simplified’ variant of U+3A18, but not the reverse. In addition, in the case of dotAsia, the variant label is blocked while in CLGR13, the preferred variant label is allocatable.

Obviously, the ‘simp’ mapping is not an actual simplified mapping but a mechanism to make U+64E4 a variant of U+3A18 in some circumstances. This looks cumbersome to our CJK expert.

**Conclusion**: The dotAsia definition is preferred by our CJK expert.【To be discussed 】

The code point U+4C7D was included in CLGR13 because of its membership in IICORE HKSCS and the code point U+4C9D was added because of its GS reference (Singapore source). In the CLGR13 (1st), U+4C7D and U+4C9D are part of a larger variant set. In dotAsia (2nd), U+4C7D and U+4C9D form a smaller variant set together with compatible values; similarly, the pair U+9BE7 and U+9CB3 forms another variant set with compatible values.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 4C7D | 䱽 | 4C7D | 䱽 | ≡ | r-trad |  | identity |
| 4C7D | 䱽 | 4C9D | 䲝 | **→** | simp |  |  |
| **←** | trad |  |  |
| 4C7D | 䱽 | 9BE7 | 鯧 | ↔ | blocked |  |  |
| 4C7D | 䱽 | 9CB3 | 鲳 | ↔ | blocked |  |  |
| 4C9D | 䲝 | 4C9D | 䲝 | ≡ | r-simp |  | identity |
| 4C9D | 䲝 | 9BE7 | 鯧 | ↔ | blocked |  |  |
| 4C9D | 䲝 | 9CB3 | 鲳 | ↔ | blocked |  |  |
| 9BE7 | 鯧 | 9BE7 | 鯧 | ≡ | r-trad |  | identity |
| 9BE7 | 鯧 | 9CB3 | 鲳 | **→** | simp |  |  |
| **←** | trad |  |  |
| 9CB3 | 鲳 | 9CB3 | 鲳 | ≡ | r-simp |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 4C7D | 䱽 | 4C7D | 䱽 | ≡ | r-trad |  | identity |
| 4C7D | 䱽 | 4C9D | 䲝 | **→** | simp |  |  |
| **←** | trad |  |  |
| 4C9D | 䲝 | 4C9D | 䲝 | ≡ | r-simp |  | identity |

Other dotAsia variant set:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 9BE7 | 鯧 | 9BE7 | 鯧 | ≡ | r-trad |  | identity |
| 9BE7 | 鯧 | 9CB3 | 鲳 | **→** | simp |  |  |
| **←** | trad |  |  |
| 9CB3 | 鲳 | 9CB3 | 鲳 | ≡ | r-simp |  | identity |

The Unihan entries for U+47CD, U+4C9D, U+9BE7 and U+9CB3 agree with having a single variant set as in CLGR13. However, the two base traditional characters correspond to two different fishes in Hong Kong and Cantonese culture. Therefore, having them blocking each other is not ideal. Given that U+9BE7 is slightly preferred over U+4C7F, the following variant set is suggested:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 4C7D | 䱽 | 4C7D | 䱽 | ≡ | r-neither |  | identity |
| 4C7D | 䱽 | 4C9D | 䲝 | **→** | simp |  |  |
| **←** | trad |  |  |
| 4C7D | 䱽 | 9BE7 | 鯧 | **→** | trad |  |  |
| **←** | simp |
| 4C7D | 䱽 | 9CB3 | 鲳 | ↔ | blocked |  |  |
| 4C9D | 䲝 | 4C9D | 䲝 | ≡ | r-simp |  | identity |
| 4C9D | 䲝 | 9BE7 | 鯧 | ↔ | blocked |  |  |
| 4C9D | 䲝 | 9CB3 | 鲳 | ↔ | blocked |  |  |
| 9BE7 | 鯧 | 9BE7 | 鯧 | ≡ | r-trad |  | identity |
| 9BE7 | 鯧 | 9CB3 | 鲳 | **→** | simp |  |  |
| **←** | trad |  |  |
| 9CB3 | 鲳 | 9CB3 | 鲳 | ≡ | r-simp |  | identity |

**Conclusion**: IP is requesting CGP to review the IP suggestion. 【To be disscused】

The code point U+701E was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between dotAsia (table follows) where it is part of the variant set U+51C0, U+51C8, U+6D44, U+6DE8, and U+701E. In CLGR13, U+701E is a singleton reflexive variant of type ‘r-both.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 51C0 | 净 | 51C0 | 净 | ≡ | r-simp |  | identity |
| 51C0 | 净 | 51C8 | 凈 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 51C0 | 净 | 6D44 | 浄 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 51C0 | 净 | 6DE8 | 淨 | **→** | trad |  |  |
| **←** | simp |  |  |
| 51C0 | 净 | 701E | 瀞 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 51C8 | 凈 | 51C8 | 凈 | ≡ | r-trad |  | identity |
| 51C8 | 凈 | 6D44 | 浄 | ↔ | blocked |  |  |
| 51C8 | 凈 | 6DE8 | 淨 | ↔ | blocked |  |  |
| 51C8 | 凈 | 701E | 瀞 | ↔ | blocked |  |  |
| 6D44 | 浄 | 6D44 | 浄 | ≡ | r-neither |  | identity |
| 6D44 | 浄 | 6DE8 | 淨 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 6D44 | 浄 | 701E | 瀞 | ↔ | blocked |  |  |
| 6DE8 | 淨 | 6DE8 | 淨 | ≡ | r-trad |  | identity |
| 6DE8 | 淨 | 701E | 瀞 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 701E | 瀞 | 701E | 瀞 | ≡ | r-neither |  | identity |

Unihan has variant relationship between U+51C0, U+51C8, U+6D44, and U+6DE8, but not U+701E. According to our CJK expert, U+701E has at least two meanings and two pronunciations and in some HK context can be a variant of the other characters.

**Conclusion**: Our CJK expert strongly prefers the dotAsia definition.【To be discussed 】

The code point U+8420 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between dotAsia (table follows) and CLGR13 (where it is a singleton reflexive variant of ‘r-both’).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 840C | 萌 | 840C | 萌 | ≡ | r-both |  | identity |
| 840C | 萌 | 8420 | 萠 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 840C | 萌 | 8544 | 蕄 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 8420 | 萠 | 8420 | 萠 | ≡ | r-trad |  | identity |
| 8420 | 萠 | 8544 | 蕄 | ↔ | blocked |  |  |
| 8544 | 蕄 | 8544 | 蕄 | ≡ | r-trad |  | identity |

In Unihan U+8420 is a zVariant of U+U+840C, therefore some variant mapping is expected (as done in the dotAsia table). The dotAsia table seems preferable because in HK context, the similarity between U+840C and U+8420 could be used for phishing purposes.

**Conclusion**: Our CJK expert strongly prefers the dotAsia definition.【To be discussed 】

## Variant sets corresponding to visual exchangeability

In section 7 of the Proposal CGP has stated that ‘*The Chinese language and script community regard only “semantically identical characters” as exchangeable variants*’. Because the CLGR13 variant sets already include members that are not cognate (in other terms, not semantically related), it is not surprising that the list of characters provided in the table part of section 7 of the proposal contains both cognate and non-cognate pairs. The part of the list pertinent to CLGR13 is repeated here (the table contains Unicode confusable, with the yellow section containing source and target glyph not part of the same variant set, while in the grey they are in the same variant set; 2 lines were added, see notes below):

|  |  |  |  |
| --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |
| 53E3 | 口 | 56D7 | 囗 |
| 571F | 土 | 58EB | 士 |
| 58AB | 墫 | 58FF | 壿 |
| 5B0E[[1]](#footnote-0) | 嬎 | 5B14 | 嬔 |
| 676E | 杮 | 67FF | 柿 |
| 8D7F | 赿 | 8D86 | 趆 |
| 9E42 | 鹂 | 9E43 | 鹃 |
| 3BA3 | 㮣 | 69E9 | 槩 |
| 4443 | 䑃 | 6726 | 朦 |
| 43D3[[2]](#footnote-1) | 䏓 | 670A | 朊 |
| 4F75 | 併 | 5002 | 倂 |
| 5024 | 値 | 503C | 值 |
| 5553 | 啓 | 555F | 啟 |
| 5861 | 塡 | 586B | 填 |
| 5AAF | 媯 | 5B00 | 嬀 |
| 5E21 | 帡 | 5E32 | 帲 |
| 6236 | 戶 | 6238 | 戸 |
| 665A | 晚 | 6669 | 晩 |
| 670F | 朏 | 80D0 | 胐 |
| 6710 | 朐 | 80CA | 胊 |
| 6713 | 朓 | 8101 | 脁 |
| 6718 | 朘 | 8127 | 脧 |
| 6723 | 朣 | 81A7 | 膧 |
| 699D | 榝 | 6A27 | 樧 |
| 6E88 | 溈 | 6F59 | 潙 |
| 7D55 | 絕 | 7D76 | 絶 |
| 80FC | 胼 | 8141 | 腁 |
| 848D | 蒍 | 853F | 蔿 |
| 8C5C | 豜 | 8C63 | 豣 |
| 8EFF | 軿 | 8F27 | 輧 |
| 90CE | 郎 | 90DE | 郞 |
| 93AD | 鎭 | 93AE | 鎮 |
| 96B7 | 隷 | 96B8 | 隸 |
| 9ED1 | 黑 | 9ED2 | 黒 |

.

Of the sets highlighted in yellow above (not part of CLGR13) the first two

|  |  |  |  |
| --- | --- | --- | --- |
| 53E3 | 口 | 56D7 | 囗 |
| 571F | 土 | 58EB | 士 |

while superficially similar, are generally well known as visually similar but do not qualify as exchangeable.

To facilitate comparison, the following yellow highlighted pairs are shown again with a red circle showing the differences:

|  |  |  |  |
| --- | --- | --- | --- |
| 58AB | 墫 | 58FF | 壿 |
| 676E | 杮 | 67FF | 柿 |
| 8D7F | 赿 | 8D86 | 趆 |

Note that that table also contains this case of a visual exchangeable pair, which is part of dotAsia:

|  |  |  |  |
| --- | --- | --- | --- |
| 5B0E | 嬎 | 5B14 | 嬔 |

In section 7 of the proposal, the CGP, while mentioning that ‘*Chinese community members disputed the concept of visual variants*’, has suggested the following dispositions:

*According to the discussion with IP in ICANN’64 and further feedback from IP, CGP would tentatively handle the 6 Han-Han pairs as below:*

*Two non-modern used pairs will be treated as visual identical variants*

*-- 676E*杮 *& 67FF*柿*、8D7F*赿 *& 8D86*趆

*Four pairs will be treated as unrelated singletons*

*-- 571F*土 *& 58EB*士*、9E42*鹂 *& 9E43*鹃*、53E3*口 *& 56D7* 囗*、58AB*墫 *& 58FF*壿

*One new pair not in the confusable list will be imported from dotAsia as visual identical variants*

*--5B0E* 嬎 *& 5B14* 嬔

The IP agrees with the CGP on these dispositions.

Based on the above, the variant set {U+8D7F, U+8D86} could be defined as follows (assuming U+8D86 is more common than U+8D7F)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 8D7F | 赿 | 8D7F | 赿 | ≡ | r-neither |  | identity |
| 8D7F | 赿 | 8D86 | 趆 | **→** | both |  |  |
| **←** | blocked |
| 8D86 | 趆 | 8D86 | 趆 | ≡ | r-both |  | identity |

The variant set {U+676E, U+67BE, U+67FF} resulting of incorporating U+676E into the pre-existing variant set {U+67BE, U+67FF} could be defined as follows: (assuming U+67FF is more common than U+676E)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 676E | 杮 | 676E | 杮 | ≡ | r-neither |  | identity |
| 676E | 杮 | 67BE | 枾 | ↔ | blocked |  |  |
| 676E | 杮 | 67FF | 柿 | **→** | both |  |  |
| **←** | blocked |
| 67BE | 枾 | 67BE | 枾 | ≡ | r-neither |  | identity |
| 67BE | 枾 | 67FF | 柿 | **→** | both |  |  |
| **←** | blocked |  |  |
| 67FF | 柿 | 67FF | 柿 | ≡ | r-both |  | identity |

The pair (U+5B0E 嬎, U+5B14 嬔) part of dotAsia could be added as follows (this is also part of section 4.2.3.1):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5B0E | 嬎 | 5B0E | 嬎 | ≡ | r-neither |  | identity |
| 5B0E | 嬎 | 5B14 | 嬔 | **→** | both |  |  |
| **←** | blocked |
| 5B14 | 嬔 | 5B14 | 嬔 | ≡ | r-both |  | identity |

Following the discussion held at ICAAN 64 meeting, IP is still convinced that the additional three variant sets should be added to CLGR (noting that one of them is already part of dotAsia).

**IP Recommendation**

* IP recommends adding the 3 variant sets as defined above in a new version of CLGR.

More considerations concerning policy about visual variants are provided in section 9.

# Issues with XML files

IP found the following issues in the XML files provided in CLGR13. (remarks in bold have been made several times, it would be good to act on them so that IP does not have to fix the XML files every time)

Main CLGR13 XML file:

* **Line 1 is using an incorrect <?xml statement, it should be: <?xml version="1.0"?>**
* **Line 59 (or thereabouts) has an un-escaped "&" - that should be &amp; instead.**
* **Line 61 (or thereabouts) is missing a closing tag </references>.**
* **Line 64 has an erroneous xml statement: ‘<?xml version="1.4" encoding="UTF-8"?>’**
* The <description> section is too terse. While the intent is not to duplicate the [Proposal] document, it should include the essential tenets of the proposal, in term of repertoire definition, variant set definition, rules, etc. The goal is to make the XML file (and its automatically HTML derived file) a self-sufficient document for implementors of the LGR.
* For example, all the variant types should be listed with an explanation of their meanings; all CLGR-specific “actions” should be listed with an explanation, and a concise explanation of the effect on allocatable variant labels provided.
* The IP strongly recommends that the CGP consult the LGR-3 XML files currently in final review as examples, and for guidance on structure and outline.

**IP Recommendation**

* **Please fix XML files according to recommendations above.**

# Issues with the Proposal document

## General

The document still contains some sections for which a detailed feedback has been provided earlier and no action has been taken. The feedback is repeated here again with more suggestion on how to fix it.

## Specific feedback concerning CLGR13 section about pre-integration

The CLGR13 [proposal] includes the following text:

***6.2.2.1 Principle and Framework***

*A coordination mechanism among three parties is needed to realize unified Chinese script generation rules in the DNS root zone. During the CDNC meeting in Shanghai (May, 2014), the IP proposed the basic principles of the coordination scheme:*

* *Each CJK panel creates an LGR and each LGR includes a repertoire and variants.*
* *If an LGR includes Han characters, the variant mappings must agree for all three panels.*
* *The variant types may be different (blocked or allocatable), so that the variant types do not have to agree as across LGRs.*

*Based on the principles above, the CGP, JGP and KGP started coordination work at the IETF Dallas meeting in 2015, trying to define a unified variant-mapping table for Chinese scripts, then define each party’s variant types/sub-types (e.g., allocatable or blocked) for characters contained in this table. According to the consensus at the IETF Dallas meeting, the JGP initiated work called "CJK Integration Procedure" as follows:*

*Step 1: Each CJK GP generates its own LGR (hereinafter, LGR-alpha)*

*Step 2: CJK GPs collectively generate a merged table of each LGR-alpha (hereinafter, LGR-M)*

*Step3: Each CJK GP extracts its original repertoire with integrated variants from LGR-M.*

*Step 4: Each CJK GP adds "Out of Repertoire" code points for symmetry.*

*Step 5: Each CJK GP merges WLE in LGR-alpha into one.*

*Step 6: Each CJK GP* generates an integrated LGR (hereinafter, LGR-beta).

After further analysis, the IP has determined that there is a need for further clarification in the text above, as it exhibits apparent contradiction to the process described in the [Procedure]:

1. While any LGR including Han characters will have variant mappings that will be integrated and therefore in ‘agreement’ among these LGRs, it is not a requirement of the GPS to produce such ‘integrated’ LGRs.
2. What is required is for GPs to provide explicit variant type definitions for any variant mappings inherited via integration. For that purpose, it is beneficial for GPs to share their respective LGRs, in order to compare their variant sets and determine expected effects of the integration.
3. CJK GPs do not generate integrated LGRs. It is the role of the Integration Panel to do so. According to the coordination process, the CJK GPS are expected to produce LGRs that are easier to integrate by the IP by harmonizing repertoire and variant sets. The process done by the CJK GPs was in fact called ‘pre-integration’ which is an acceptable qualifier.
4. The steps 2 to 6 do not correspond to process steps expected to be conducted by the GPs. While nothing prevents the GPs from carrying out a ‘virtual’ integration, it is not a requirement. For example, consider the step 4 described above ‘The CJK GPS adds “out-of-repertoire” code points for symmetry’: this has not been consistently done by all the GPs. The only time an out-of-repertoire variant is strictly required is if it is the only way to define a cross-repertoire variant (because the variant is not in-repertoire for any LGR). See [OutofRepertoireVariants] for further details.
5. For variants that are in-repertoire in one LGR, but cross-repertoire in another LGR, the IP can mechanically impose those variants required for symmetry and transitivity. If a GP does the same thing, this can serve as a double check — assuming it is done correctly.
6. The pre-integration done between CGP and KGP would introduce significant modification to the Chinese LGRs resulting in significant differences between planned practice for Chinese IDN (such as CDNC-2015) and the Root Zone LGR. These changes would appear to go against other tenets of the [Procedure], especially the ‘Stability Principle’. One would expect that everything that CDNC-2015 considers to be exchangeable variants would also be exchangeable variants in the root zone. In addition, the ‘Conservatism Principle’ would strongly argue against a reduction of blocked variants and the weakening of the protections due to removing some of these variants sets would appear to contrast with the “Contextual Safety Principle” and would potentially make the root zone less safe than lower level zones.
7. In the interest of stability security, the CJK GPs are encouraged to preserve their original variant sets as much as possible, and to focus on defining the LGR-specific variant types for shared variant mappings.

For example, in the case of two code points that are considered exchangeable in the context of Chinese users but distinct in Korean Hanja, the Conservatism and Security principles would allow preserving the allocatable disposition in the Chinese LGR and a blocked disposition in the Korean LGR. On the first come first served basis, a Korean user could still apply for either of the labels containing these code points if they act prior to an application for either that label or its variant (whether by a Chinese user or another Korean user). Conversely, a Chinese user may act first and apply for both the label and its allocatable variant.

The fact that in some scenarios some labels may not be delegated (because they are blocked variants of a delegated label) is immaterial. In Section B.4.1, the Procedure states (emphasis added):

*“[This may] produce labels that would violate the linguistic criteria for being considered true variant and may also result in the generation of extra blocked variants that lead to the exclusion of other possibly useful labels. It is nevertheless appropriate in the root zone, where the* ***goal is not to maximize the number of possible labels*** *but to minimize the confusion possible in a shared environment supporting heterogeneous linguistic communities.”*

Blocking labels that are considered the same in a Chinese context is superior to the removal of the variant sets (as currently proposed through the current ‘pre-integration’), which would allow separate Chinese entities to delegate labels that would be perceived as exchangeable by most users of the root zone familiar with the Han script.

A further consideration is that changing hundreds of variant sets (as done in CLGR context) introduces a major risk of mistakes that are hard to detect. Many variant sets have multiple code point members (up to 8 members), and the removal of one or more members would require a complex analysis of the effect on the various remaining mappings.

**IP Recommendation**

* Rewrite section ‘6.2.2 CJK Coordination, taking into account IP feedback provided above.
* At minimum, a new section should avoid the misleading implication that the statement: ‘*If an LGR includes Han characters, the variant mappings must agree for all three panels*’ means that variant mappings are to be agreed upon; instead, they simply have to be the union of all individual LGRs’ mappings.
* In addition it should not include steps 2 to 6 in 6.2.2.1 which is not an agreed part of the integration procedure (again CJK GPs are not in charge of integration, this is the role of the Integration Panel).
* The section 6.2.2.2 ‘C-J Coordination’ is not controversial and could stay in the [Proposal]. On the other hand the section 6.2.2.3 ‘C-K Coordination’ while describing correctly the process engaged in the past 2 years, is extremely misleading (for the reasons exposed above) and should be removed from the proposal, unless completely rewritten.

## Mechanism for reducing multiple allocatable labels

The proposed mechanism in Section 6.3 of [Proposal] is a feasible approach, at the expense of additional complexity in the WLE rules. While the text has been improved from last version, the IP found some discrepancies and need for further explanation.

In the [Proposal] text in section 6.3.1, the examples given do not correspond directly to each of the categories of multiple allocatable mappings. It would greatly help the clarity of the proposal if the examples, along with the assigned types, correspond to the categories. Section 6.3.1 should be a full introduction of the **formal** definitions made in section 6.3.2.

The table identifying the new mapping sub-type has the following issues:

* The column ‘Type’ is incorrect because it lists only one of the possible type but should either contain all possible dispositions (a better header for the column would be ‘Disposition’) or removed altogether because the comment already gives the information. 【done】
* The table could add some explanation of how these various reflexive types work. The following is a guess by IP based on the related actions (this feedback was already provided for CLGR11 and CLGR12): 【TBD】
  + r-both-ms indicates that for a given code point, its reflexive type is inherently r-both, but there is at least another ‘simp’ type (or other simplified types), and therefore it is preferred in a traditional context. Therefore, it is to be treated as a ‘r-trad’.
  + r-both-mt indicates that for a given code point, its reflexive type is inherently r-both, but there is at least another ‘trad’ type (or other traditional types), and therefore it is preferred in a simplified context. Therefore, it is to be treated as a ‘r-simp’.
  + r-simp-m indicates that for a given code point, its reflexive type is inherently r-simp, but there is at least another ‘simp’ type (or other simplified types), along with another ‘trad’ type and therefore it is never preferred in any variant labels. Therefore, it is to be treated as a ‘r-neither’. Given that it is only used once, one may wonder why that occurrence was just not encoded with a ‘r-neither’ to facilitate comprehension. Note that there is no ‘r-trad-m’ defined although it could also exist in principle (but would also be equivalent to a ‘r-neither’).
* ALLOCATABLE is misspelt as ‘ALLOATABLE.【done】

The IP found the following issues in the implementation of the XML as implemented following Section 6 of the proposal:

* There are nine occurrences where the order of code points for ‘trad-1’ and ‘trad-2’ does not respect the criterion for assigning trad-1 and trad-2 defined in the table; that is by code point order (this feedback was already provided for CLGR11 and CLGR13): 【done】
  + For code point 5364: Variant 9E75(trad-1) > Variant 6EF7(trad-2).
  + For code point 5386: Variant 6B77(trad-1) > Variant 66C6(trad-2).
  + For code point 5F53: Variant 7576(trad-1) > Variant 5679(trad-2).
  + For code point 6076: Variant 60E1(trad-1) > Variant 5641(trad-2).
  + For code point 82CF: Variant 8607(trad-1) > Variant 7526(trad-2).
  + For code point 8303: Variant 8303(trad-1) > Variant 7BC4(trad-2).
  + For code point 8D5E: Variant 8D0A(trad-1) > Variant 8B9A(trad-2).
  + For code point 94FA: Variant 92EA(trad-1) > Variant 8216(trad-2).
  + For code point 9988: Variant 994B(trad-1) > Variant 993D(trad-2).

If there is a reason for deviating from the code point order in assinging trad-1 and trad-2, the actual selection criteria should be documented and a justification given.

* The variant set {7BC4, 8303} does not have a reflexive type for U+8303. This looks like a typo and it should probably be changed to ‘r-both-mt’ like in CLGR12. 【done】
* The mapping types ‘simp-1’ and ‘simp-2’ have been misspelt ‘Simp-1’ and ‘Simp-2’ in some parts of the XML file, which creates discrepancy between the variant mapping and the action section (values in the XML are case sensitive). 【TBD】
* Three pairs of reflexive types: r-simp / r-both-mt, r-trad / r-both-ms’, r-neither / r-simp-m only occur grouped together, therefore these types are functionally equivalent. While their descriptive names reflect the way they are derived by algorithm, using a single name for their function would perhaps facilitate comprehension of the mechanism. If they are retained, some text should be added in the proposal to indicate that how both members of each pair lead to identical dispositions. 【TBD】

Based on our understanding of the mechanism, any change to the variant sets from CLGR13 requires recalculating the various ‘muted’ types for the modified variant sets. The IP has explored that path and determined that these muted types can be computed by algorithm. The only exception is ‘trad-m’ for which there is no automated mechanism to determine which code point value is to be ‘muted’ (in effect blocked) among a triplet of ‘trad’ variants. Currently there are only two mappings with trad-m type, U+53F0**→**U+98B1 and U+8499**→**U+61DE.

**IP Recommendation**

* Implement suggestions provided in the IP remarks above, or give rationale for current preference.

## Visually Identical Characters

Some feedback is already provided in section 4.4 concerning the variant sets to be added to the proposal in the XML file. However, along with these added variant sets, the [Proposal] related section (7) contains many details that seem superfluous. For example, details about possible confusion concerning the Japanese or Korean writing systems don’t apply to Han ideographs. The scope of possible visual confusion in these ideographs has been extremely limited and has resulted in only adding 3 variant sets in CLGR13 (one being already part of dotAsia).

Describing an excerpt of the confusables.txt file from Unicode is also misleading because it was not endorsed by either CGP or IP, but only used as possible input. Only a very small subset of that table was considered as possible visual variants in the context of the CLGR.

In general, this section should be reduced to describe the eventual consensus among parties (IP and CGP) with a brief description on the historic process.

**IP Recommendation**

* Implement suggestions provided in the IP remarks above.

PART III — Background and detail analysis

# Notation and explanation used in the description of the variant set differences

## General

Section 4 and section 8 describe in detail some of the differences between CLGR13 and dotAsia. Where the variants set from the two LGRs differ in their definition but are overlapping, tables listing both variant sets are shown together and the header indications whether the set from CLGR13 or dotAsia comes first or second. Differences in variant mappings are highlighted in red. Although the comparison was mostly mechanically generated, some manual editing was necessary.

## Table format for variant sets

The format of each table follows the format used in the HTML-formatted LGR tables, as used in LGR-2. The tables list each pair of variant mappings on one row. For each pair of code points, by convention, the lower code point is taken as the source of the mapping in the forward → direction and information for the reverse ← direction is usually not listed separately. The variant mappings defined in an LGR are required to be symmetric, that is, both the forward and reverse mappings must be specified.

A mapping where source and target are the same is reflexive. Variant sets consisting of only a single reflexive mapping are not shown as a set. Instead, the variant type of the mapping is listed in the Variants column of the Repertoire by Code Point table. Reflexive mappings that are part of a larger set are indicated with a “≡”.

Where the type of both forward and reverse mappings are the same, a single value is given in the Type(s) column, otherwise the types for forward and reverse mapping are given in that order, as indicated by the arrows. The same applies to any comments.

In a properly specified LGR, all members of each variant set are variants of each other, a property called transitivity. Therefore, all variant sets are disjoint.

The variant sets are presented in increasing numerical order of source code points and target code points in the set, irrespective of which code point(s) have a difference in mapping between the LGRs being compared. This is to facilitate comparison with the original XML file (or HTML transcription).

When related sets from CLGR13 and dotAsia are presented together, the CLGR13 variant is always first. In each table, the background color alternates whenever the source code point in the left column changes, thus grouping all mappings using the same source code point.

# Detailed review of variant set differences

## Review of variant sets for additional HKSCS characters beyond dotAsia – 2 sets

The following table shows variants sets modified by the addition of these 2 new HKSCS characters derived from HKIRC (shown in red in the table).

| **No** | **CLGR13 Variant set glyphs and code points** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 㧛 39DB | 㩜 3A5C | 揽63FD | 擥64E5 | 攬652C |  |  |  |
| 2 | 坛575B | 埮57EE | 墰58B0 | 墵58B5 | 壇58C7 | 壜58DC | 罈7F48 | 罎7F4E |

This variant set has one added member U+3A5C (not in dotAsia) from the HKSCS set (but part of CDNC 2018). The code point U+64E5 is part of both dotAsia and CDNC 2018. The code point U+39DB is part of both dotAsia and CLGR13 (but not CDNC 2018). This is a rather complex hybrid case which is covered by this section (4.1), and the section covering CLGR13 addition not part of CDNC 2018 (section 4.3).

The CLGR13 variant set (1st table) are mapped differently from dotAsia (2nd table). The red highlighting in both tables reflects all differences between the two LGRs. The CDNC 2018 variant set is defined in the 3rd table.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 39DB | 㧛 | 39DB | 㧛 | ≡ | r-simp |  | identity |
| 39DB | 㧛 | 3A5C | 㩜 | ↔ | blocked |  |  |
| 39DB | 㧛 | 63FD | 揽 | ↔ | blocked |  |  |
| 39DB | 㧛 | 64E5 | 擥 | **→** | trad |  |  |
| **←** | blocked |
| 39DB | 㧛 | 652C | 攬 | ↔ | blocked |  |  |
| 3A5C | 㩜 | 3A5C | 㩜 | ≡ | r-both |  | identity |
| 3A5C | 㩜 | 63FD | 揽 | ↔ | blocked |  |  |
| 3A5C | 㩜 | 64E5 | 擥 | ↔ | blocked |  |  |
| 3A5C | 㩜 | 652C | 攬 | ↔ | blocked |  |  |
| 63FD | 揽 | 63FD | 揽 | ≡ | r-simp |  | identity |
| 63FD | 揽 | 64E5 | 擥 | ↔ | blocked |  |  |
| 63FD | 揽 | 652C | 攬 | **→** | trad |  |  |
| **←** | simp |  |  |
| 64E5 | 擥 | 64E5 | 擥 | ↔ | r-both |  | identity |
| 64E5 | 擥 | 652C | 攬 | ↔ | blocked |  |  |
| 652C | 攬 | 652C | 攬 | ≡ | r-trad |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 39DB | 㧛 | 39DB | 㧛 | ≡ | r-simp |  | identity |
| 39DB | 㧛 | 63FD | 揽 | ↔ | blocked |  |  |
| 39DB | 㧛 | 64E5 | 擥 | **→** | trad |  |  |
| **←** | simp |  |  |
| 39DB | 㧛 | 652C | 攬 | ↔ | blocked |  |  |
| 63FD | 揽 | 63FD | 揽 | ≡ | r-simp |  | identity |
| 63FD | 揽 | 64E5 | 擥 | ↔ | blocked |  |  |
| 63FD | 揽 | 652C | 攬 | **→** | trad |  |  |
| **←** | simp |  |  |
| 64E5 | 擥 | 64E5 | 擥 | ≡ | r-trad |  | identity |
| 64E5 | 擥 | 652C | 攬 | ↔ | blocked |  |  |
| 652C | 攬 | 652C | 攬 | ≡ | r-trad |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 3A5C | 㩜 | 3A5C | 㩜 | ≡ | r-both |  | identity |
| 3A5C | 㩜 | 63FD | 揽 | ↔ | blocked |  |  |
| 3A5C | 㩜 | 64E5 | 擥 | ↔ | blocked |  |  |
| 3A5C | 㩜 | 652C | 攬 | ↔ | blocked |  |  |
| 63FD | 揽 | 63FD | 揽 | ≡ | r-simp |  | identity |
| 63FD | 揽 | 64E5 | 擥 | ↔ | blocked |  |  |
| 63FD | 揽 | 652C | 攬 | **→** | trad |  |  |
| **←** | simp |  |  |
| 64E5 | 擥 | 64E5 | 擥 | ≡ | r-both |  | identity |
| 64E5 | 擥 | 652C | 攬 | ↔ | blocked |  |  |
| 652C | 攬 | 652C | 攬 | ≡ | r-trad |  | identity |

Unihan indicates that U+3A5C is a semantic variant of U+652C 攬and U+64E5 擥. In addition, U+3A2B 㨫 (not in CLGR13) is listed as a simplified variant. The mapping for U+3A5C is acceptable (and corresponds to an earlier feedback from IP). IP also note that the mappings for the pair (U+64E5, U+39DB) as requested in the updated Annex D worksheet: ‘6.2.1.3-34’ now matches the variant set definition.

**Conclusion**: The CLGR13 version is acceptable.

This variant set has one added member U+58B5 (not in dotAsia.) from the HKSCS set coming from HKIRC. In addition, U+58DC part of CLGR13 (1st table) and dotAsia (2nd table) is mapped differently between the two sets. This case is a hybrid of this category (one code point: U+58B5 added not in dotAsia) and the next category (one code point part of both CDNC 2018 and dotAsia: U+58DC but treated differently covered in section 4.2).

In addition, the dotAsia variant set does not include U+57EE and U+58B0 (both are singleton reflexive variants of type ‘r-both’), while the CLGR13 table add them as ‘blocked’ variants of all other members. The CLGR13 set is an 8 members variant set versus a 5 members variant set for dotAsia.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 575B | 坛 | 575B | 坛 | ≡ | r-simp |  | identity |
| 575B | 坛 | 57EE | 埮 | ↔ | blocked |  |  |
| 575B | 坛 | 58B0 | 墰 | ↔ | blocked |  |  |
| 575B | 坛 | 58B5 | 墵 | ↔ | blocked |  |  |
| 575B | 坛 | 58C7 | 壇 | **→** | trad |  |  |
| **←** | simp |  |  |
| 575B | 坛 | 58DC | 壜 | ↔ | blocked |  |  |
| 575B | 坛 | 7F48 | 罈 | **→** | trad |  |  |
| **←** | simp |  |  |
| 575B | 坛 | 7F4E | 罎 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 57EE | 埮 | 57EE | 埮 | ≡ | r-both |  | identity |
| 57EE | 埮 | 58B0 | 墰 | ↔ | blocked |  |  |
| 57EE | 埮 | 58B5 | 墵 | ↔ | blocked |  |  |
| 57EE | 埮 | 58C7 | 壇 | ↔ | blocked |  |  |
| 57EE | 埮 | 58DC | 壜 | ↔ | blocked |  |  |
| 57EE | 埮 | 7F48 | 罈 | ↔ | blocked |  |  |
| 57EE | 埮 | 7F4E | 罎 | ↔ | blocked |  |  |
| 58B0 | 墰 | 58B0 | 墰 | ≡ | r-both |  | identity |
| 58B0 | 墰 | 58B5 | 墵 | ↔ | blocked |  |  |
| 58B0 | 墰 | 58C7 | 壇 | ↔ | blocked |  |  |
| 58B0 | 墰 | 58DC | 壜 | ↔ | blocked |  |  |
| 58B0 | 墰 | 7F48 | 罈 | ↔ | blocked |  |  |
| 58B0 | 墰 | 7F4E | 罎 | ↔ | blocked |  |  |
| 58B5 | 墵 | 58B5 | 墵 | ≡ | r-both |  | identity |
| 58B5 | 墵 | 58C7 | 壇 | ↔ | blocked |  |  |
| 58B5 | 墵 | 58DC | 壜 | ↔ | blocked |  |  |
| 58B5 | 墵 | 7F48 | 罈 | ↔ | blocked |  |  |
| 58B5 | 墵 | 7F4E | 罎 | ↔ | blocked |  |  |
| 58C7 | 壇 | 58C7 | 壇 | ≡ | r-trad |  | identity |
| 58C7 | 壇 | 58DC | 壜 | ↔ | blocked |  |  |
| 58C7 | 壇 | 7F48 | 罈 | ↔ | blocked |  |  |
| 58C7 | 壇 | 7F4E | 罎 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 58DC | 壜 | 58DC | 壜 | ≡ | r-both |  | identity |
| 58DC | 壜 | 7F48 | 罈 | ↔ | blocked |  |  |
| 58DC | 壜 | 7F4E | 罎 | ↔ | blocked |  |  |
| 7F48 | 罈 | 7F48 | 罈 | ≡ | r-trad |  | identity |
| 7F48 | 罈 | 7F4E | 罎 | ↔ | blocked |  |  |
| 7F4E | 罎 | 7F4E | 罎 | ≡ | r-neither |  | identity |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 575B | 坛 | 575B | 坛 | ≡ | r-simp |  | identity |
| 575B | 坛 | 58C7 | 壇 | **→** | trad |  |  |
| **←** | simp |  |  |
| 575B | 坛 | 58DC | 壜 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 575B | 坛 | 7F48 | 罈 | **→** | trad |  |  |
| **←** | simp |  |  |
| 575B | 坛 | 7F4E | 罎 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 58C7 | 壇 | 58C7 | 壇 | ≡ | r-trad |  | identity |
| 58C7 | 壇 | 58DC | 壜 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 58C7 | 壇 | 7F48 | 罈 | ↔ | blocked |  |  |
| 58C7 | 壇 | 7F4E | 罎 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 58DC | 壜 | 58DC | 壜 | ≡ | r-neither |  | identity |
| 58DC | 壜 | 7F48 | 罈 | ↔ | blocked |  |  |
| 58DC | 壜 | 7F4E | 罎 | ↔ | blocked |  |  |
| 7F48 | 罈 | 7F48 | 罈 | ≡ | r-trad |  | identity |
| 7F48 | 罈 | 7F4E | 罎 | ↔ | blocked |  |  |
| 7F4E | 罎 | 7F4E | 罎 | ≡ | r-neither |  | identity |

**Conclusion**: The new CLGR13 corresponds to the recommendation that was made by IP on earlier feedback and as such, the set is acceptable.

## Detailed list of differences for CDNC2018/dotAsia additions – 35 sets

The following numbered items list all the information available to the IP at this point for each of the differences found and its recommendations based on evidence and the opinion of the IP CJK expert.

The code point U+53DA was included in CLGR13 because of its membership in IICORE HKSCS but is treated differently between CLGR13 and dotAsia (table). In CLG13, both U+5047 and U+53DA are singleton reflexive variants of type ‘r-both’.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5047 | 假 | 5047 | 假 | ≡ | r-both |  | identity |
| 5047 | 假 | 53DA | 叚 | ↔ | blocked |  |  |
| 53DA | 叚 | 53DA | 叚 | ≡ | r-both |  | identity |

Unihan does not bring any clarification either way (it describes a variant relationship between U+5047 and U+4EEE part of CLGR13 but not included here).

In CDNC-2018, U+5047, U+53DA, and U+6BB5 are all reflexive singletons. In answer to IP feedback to CLGR10, CGP mentioned that the dotAsia was ‘*more reasonable*’. In previous versions of CLGR, a variant mapping with U+6BB5 段 was also suggested but has been now abandoned.

**Conclusion**: After further review by our CJK expert, the CLGR13 version is endorsed by IP.

The code point U+681E was included in CLGR13 because of its membership in IICORE HKSCS . In dotAsia (table), U+520A, U+520B, and U+681E have a variant relationship. In CLGR13 and CDNC-2018, U+681E is a singleton reflexive variant of type ‘r-both’.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 520A | 刊 | 520A | 刊 | ≡ | r-both |  | identity |
| 520A | 刊 | 520B | 刋 | **→** | blocked |  |  |
| **←** | both |  |  |
| 520A | 刊 | 681E | 栞 | **→** | blocked |  |  |
| **←** | both |  |  |
| 520B | 刋 | 520B | 刋 | ≡ | r-neither |  | identity |
| 520B | 刋 | 681E | 栞 | ↔ | blocked |  |  |
| 681E | 栞 | 681E | 栞 | ≡ | r-neither |  | identity |

Unihan has a variant relationship between U+520A and U+520B, but not U+681E. Code points U+520A and U+681E are part of KLGR (but not part of CK pre-integration).

**Conclusion**: Based on Unihan and additional review, the CLGR13 set is acceptable.

The code point U+52A4 was included in CLGR13 because of its membership in IICORE HKSCS. In dotAsia U+52A4, U+52B2, and U+52C1 (table follows) have a variant relationship. In CLGR13, U+52A4 is a singleton reflexive variant of type ‘r-both’, other mappings are unchanged.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 52A4 | 劤 | 52A4 | 劤 | ≡ | r-both |  | identity |
| 52A4 | 劤 | 52B2 | 劲 | ↔ | blocked |  |  |
| 52A4 | 劤 | 52C1 | 勁 | ↔ | blocked |  |  |
| 52B2 | 劲 | 52B2 | 劲 | ≡ | r-simp |  | identity |
| 52B2 | 劲 | 52C1 | 勁 | **→** | trad |  |  |
| **←** | simp |  |  |
| 52C1 | 勁 | 52C1 | 勁 | ≡ | r-trad |  | identity |

Unihan has a variant relationship between U+52B2, U+52C1, and U+4FD3 (not part of the set above), but not U+52A4.

**Conclusion**: Based on Unihan and additional review, the CLGR13 set is acceptable.

The code point U+767A was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (table follows) and dotAsia (not shown). The only difference is the type of reflexive mapping for U+767A, ‘r-neither’ for CLGR13 and ‘r-trad’ for dotAsia.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 53D1 | 发 | 53D1 | 发 | ≡ | r-simp |  | identity |
| 53D1 | 发 | 5F42 | 彂 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 53D1 | 发 | 767A | 発 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 53D1 | 发 | 767C | 發 | **→** | trad |  |  |
| **←** | simp |  |  |
| 53D1 | 发 | 9AEA | 髪 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 53D1 | 发 | 9AEE | 髮 | **→** | trad |  |  |
| **←** | simp |  |  |
| 5F42 | 彂 | 5F42 | 彂 | ≡ | r-neither |  | identity |
| 5F42 | 彂 | 767A | 発 | ↔ | blocked |  |  |
| 5F42 | 彂 | 767C | 發 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 5F42 | 彂 | 9AEA | 髪 | ↔ | blocked |  |  |
| 5F42 | 彂 | 9AEE | 髮 | ↔ | blocked |  |  |
| 767A | 発 | 767A | 発 | ≡ | r-neither |  | identity |
| 767A | 発 | 767C | 發 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 767A | 発 | 9AEA | 髪 | ↔ | blocked |  |  |
| 767A | 発 | 9AEE | 髮 | ↔ | blocked |  |  |
| 767C | 發 | 767C | 發 | ≡ | r-trad |  | identity |
| 767C | 發 | 9AEA | 髪 | ↔ | blocked |  |  |
| 767C | 發 | 9AEE | 髮 | ↔ | blocked |  |  |
| 9AEA | 髪 | 9AEA | 髪 | ≡ | r-neither |  | identity |
| 9AEA | 髪 | 9AEE | 髮 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 9AEE | 髮 | 9AEE | 髮 | ≡ | r-trad |  | identity |

Unihan tends to support the CLGR13 mapping, because only U+767C and U+9AEA are traditional mapping for U+53D1. Furthermore, U+767A is shown as a kZVariant of U+767C in Unihan. Unihan has a variant relationship between all code points.

**Conclusion**: Based on Unihan and additional review, the CLGR13 set is acceptable.

The code point U+8117 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 543B | 吻 | 543B | 吻 | ≡ | r-both |  | identity |
| 543B | 吻 | 5445 | 呅 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 543B | 吻 | 5461 | 呡 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 543B | 吻 | 8117 | 脗 | **→** | blocked |  |  |
| **←** | both |  |  |
| 5445 | 呅 | 5445 | 呅 | ≡ | r-trad |  | identity |
| 5445 | 呅 | 5461 | 呡 | ↔ | blocked |  |  |
| 5445 | 呅 | 8117 | 脗 | ↔ | blocked |  |  |
| 5461 | 呡 | 5461 | 呡 | ≡ | r-trad |  | identity |
| 5461 | 呡 | 8117 | 脗 | ↔ | blocked |  |  |
| 8117 | 脗 | 8117 | 脗 | ≡ | r-neither |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 543B | 吻 | 543B | 吻 | ≡ | r-both |  | identity |
| 543B | 吻 | 5445 | 呅 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 543B | 吻 | 5461 | 呡 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 543B | 吻 | 8117 | 脗 | ↔ | blocked |  |  |
| 5445 | 呅 | 5445 | 呅 | ≡ | r-trad |  | identity |
| 5445 | 呅 | 5461 | 呡 | ↔ | blocked |  |  |
| 5445 | 呅 | 8117 | 脗 | ↔ | blocked |  |  |
| 5461 | 呡 | 5461 | 呡 | ≡ | r-trad |  | identity |
| 5461 | 呡 | 8117 | 脗 | ↔ | blocked |  |  |
| 8117 | 脗 | 8117 | 脗 | ≡ | r-both |  | identity |

Unihan has a mapping between U+543B and U+5461 (and U+80B3 not part of this variant set), but not U+5445 or U+8117.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+8597 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 56ED | 园 | 56ED | 园 | ≡ | r-simp |  | identity |
| 56ED | 园 | 5712 | 園 | **→** | trad |  |  |
| **←** | simp |  |  |
| 56ED | 园 | 8597 | 薗 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5712 | 園 | 5712 | 園 | ≡ | r-trad |  | identity |
| 5712 | 園 | 8597 | 薗 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 8597 | 薗 | 8597 | 薗 | ≡ | r-neither |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 56ED | 园 | 56ED | 园 | ≡ | r-simp |  | identity |
| 56ED | 园 | 5712 | 園 | **→** | trad |  |  |
| **←** | simp |  |  |
| 56ED | 园 | 8597 | 薗 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5712 | 園 | 5712 | 園 | ≡ | r-trad |  | identity |
| 5712 | 園 | 8597 | 薗 | ↔ | blocked |  |  |
| 8597 | 薗 | 8597 | 薗 | ≡ | r-trad |  | identity |

Unihan has a variant relationship between U+56ED and U+5712 (and a kSemanticVariant mapping to U+5313 not in this variant set), but no mapping to U+8597.

**Conclusion**: Because CDNC-2018 has the same definition as CLGR13, and based on further review, the CLGR13 set is acceptable.

The code points 坛575B, 埮57EE, 墰58B0, 墵58B5, 壇58C7, 壜58DC, 罈7F48, 罎7F4E are part of a variant set discussed earlier in 4.1.

The code point U+57D7 was included in CLGR13 because of its membership in IICORE HKSCS but is treated differently between CLGR13 (table follows) and dotAsia (where it is a singleton reflexive variant of ‘r-both’ along with U+57E0).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 57D7 | 埗 | 57D7 | 埗 | ≡ | r-both |  | identity |
| 57D7 | 埗 | 57E0 | 埠 | ↔ | blocked |  |  |
| 57E0 | 埠 | 57E0 | 埠 | ≡ | r-both |  | identity |

Unihan kDefinition field for U+57D7 and U+57E0 indicates some similarity.

**Conclusion**: As such, the CLGR13 definition is acceptable and also agreed by our CJK expert.

The code points U+5B0E and U+5B14 form their own variant set in dotAsia. In CLGR13, these two code points are singletons with reflexive variant of type ‘r-both’. See 4.2.3.1 for details.

The code point U+5BD7 was included in CLGR13 because of its membership in IICORE HKSCS but is treated differently between CLGR13 (1st) and dotAsia (2nd). In DotAsia, U+752F is a singleton reflexive variant of type ‘r-both’. In CLGR13, U+5BD7 has different variant mappings, and U+752F is part of this variant set. See 4.2.3.2 for details.

The code point U+73E4 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (table follows) and dotAsia (where it is a singleton reflexive variant of type ‘r-both’).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5B9D | 宝 | 5B9D | 宝 | ≡ | r-simp |  | identity |
| 5B9D | 宝 | 5BF3 | 寳 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5B9D | 宝 | 5BF6 | 寶 | **→** | trad |  |  |
| **←** | simp |  |  |
| 5B9D | 宝 | 73E4 | 珤 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5BF3 | 寳 | 5BF3 | 寳 | ≡ | r-neither |  | identity |
| 5BF3 | 寳 | 5BF6 | 寶 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 5BF3 | 寳 | 73E4 | 珤 | ↔ | blocked |  |  |
| 5BF6 | 寶 | 5BF6 | 寶 | ≡ | r-trad |  | identity |
| 5BF6 | 寶 | 73E4 | 珤 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 73E4 | 珤 | 73E4 | 珤 | ≡ | r-neither |  | identity |

Unihan has a variant set for U+5B9D, U+5BF3, and U+5BF6, not including U+73E4. This supports the dotAsia mapping, not the CLGR13 set.

**Conclusion**: However, because CDNC-2018 has the same definition as CLGR13, this is acceptable. This was also endorsed by our CJK expert.

The code point U+7B92 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5E1A | 帚 | 5E1A | 帚 | ≡ | r-both |  | identity |
| 5E1A | 帚 | 7B92 | 箒 | **→** | blocked |  |  |
| **←** | both |  |  |
| 5E1A | 帚 | 83F7 | 菷 | **→** | blocked |  |  |
| **←** | both |  |  |
| 7B92 | 箒 | 7B92 | 箒 | ≡ | r-neither |  | identity |
| 7B92 | 箒 | 83F7 | 菷 | ↔ | blocked |  |  |
| 83F7 | 菷 | 83F7 | 菷 | ≡ | r-neither |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5E1A | 帚 | 5E1A | 帚 | ≡ | r-both |  | identity |
| 5E1A | 帚 | 7B92 | 箒 | ↔ | blocked |  |  |
| 5E1A | 帚 | 83F7 | 菷 | **→** | blocked |  |  |
| **←** | both |  |  |
| 7B92 | 箒 | 7B92 | 箒 | ≡ | r-both |  | identity |
| 7B92 | 箒 | 83F7 | 菷 | ↔ | blocked |  |  |
| 83F7 | 菷 | 83F7 | 菷 | ≡ | r-neither |  | identity |

Unihan has a variant for all these code points. It indicates that U+5E1A is a semantic variant of U+7B92. Therefore, the correlation is established, but not the type of preferred mapping or the reason why there is a difference between CLGR13 and dotAsia.

**Conclusion**: Because CDNC-2018 has the same definition as CLGR13, this is acceptable. This was also endorsed by our CJK expert.

The code point U+617D was included in CLGR13 because of its membership in IICORE HKSCS but with different mapping between CLGR13 (1st) and dotAsia (2nd). In one case, it is associated with U+617C, in the other with U+621A.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 617C | 慼 | 617C | 慼 | ≡ | r-both |  | identity |
| 617C | 慼 | 617D | 慽 | ↔ | blocked |  |  |
| 617D | 慽 | 617D | 慽 | ≡ | r-both |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 617D | 慽 | 617D | 慽 | ≡ | r-both |  | identity |
| 617D | 慽 | 621A | 戚 | ↔ | blocked |  |  |
| 621A | 戚 | 621A | 戚 | ≡ | r-both |  | identity |

In Unihan U+617C is a kSemanticVariant of U+617D and vice versa. There is no such association with U+621A.

**Conclusion**: In consequence of the above, the CLGR13 mapping is preferable.

The code point U+64E1 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd). Note that U+64E1 is mostly used in a Korean context.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 62AC | 抬 | 62AC | 抬 | ≡ | r-both |  | identity |
| 62AC | 抬 | 64E1 | 擡 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 64E1 | 擡 | 64E1 | 擡 | ≡ | r-trad |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 62AC | 抬 | 62AC | 抬 | ≡ | r-both |  | identity |
| 62AC | 抬 | 64E1 | 擡 | ↔ | blocked |  |  |
| 64E1 | 擡 | 64E1 | 擡 | ≡ | r-both |  | identity |

The correlation between U+62AC and U+64E1 is confirmed by Unihan (kSemanticVariant), but the difference in mappings cannot be clarified.

**Conclusion**: Because CDNC-2018 has the same definition as CLGR13, this is acceptable and was also endorsed by our CJK expert.

In these tables, the types of the mappings between U+67A3 and U+6806 or U+68D7 are reversed between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 67A3 | 枣 | 67A3 | 枣 | ≡ | r-simp |  | identity |
| 67A3 | 枣 | 6806 | 栆 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 67A3 | 枣 | 68D7 | 棗 | **→** | trad |  |  |
| **←** | simp |  |  |
| 6806 | 栆 | 6806 | 栆 | ≡ | r-trad |  | identity |
| 6806 | 栆 | 68D7 | 棗 | ↔ | blocked |  |  |
| 68D7 | 棗 | 68D7 | 棗 | ≡ | r-trad |  | identity |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 67A3 | 枣 | 67A3 | 枣 | ≡ | r-simp |  | identity |
| 67A3 | 枣 | 6806 | 栆 | **→** | trad |  |  |
| **←** | simp |  |  |
| 67A3 | 枣 | 68D7 | 棗 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 6806 | 栆 | 6806 | 栆 | ≡ | r-trad |  | identity |
| 6806 | 栆 | 68D7 | 棗 | ↔ | blocked |  |  |
| 68D7 | 棗 | 68D7 | 棗 | ≡ | r-trad |  | identity |

Unihan confirms the CLGR13 option (U+68D7 kTraditionalVariant for U+67A3). This is confirmed by CDNC-2018 and therefore acceptable. This was also endorsed by our CJK expert.

**Conclusion**: Based on Unihan and further review, the CLGR13 set is acceptable.

In CLGR13 U+68C5 maps to U+67C4 with a type of ‘both’ as shown in the following table. This is different from both .cn and dotAsia where U+68C5 maps to itself as ‘r-both’ type and where there is no variant relation between these code points (i.e. they are singletons).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 67C4 | 柄 | 67C4 | 柄 | ≡ | r-both |  | identity |
| 67C4 | 柄 | 68C5 | 棅 | **→** | blocked |  |  |
| **←** | both |  |  |
| 68C5 | 棅 | 68C5 | 棅 | ≡ | r-neither |  | identity |

Unihan defines kSemanticVariant mapping between U+67C4 and U+68C5, which could support the CLGR13 mapping.

**Conclusion**: This is confirmed by CDNC-2018, and after further review is acceptable.

The code points U+6900 and U+76CC were included in CLGR13 because of their IICORE HKSCS property but are treated differently between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 6900 | 椀 | 6900 | 椀 | ≡ | r-simp |  | identity |
| 6900 | 椀 | 76CC | 盌 | ↔ | blocked |  |  |
| 6900 | 椀 | 7897 | 碗 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 6900 | 椀 | 92FA | 鋺 | ↔ | blocked |  |  |
| 76CC | 盌 | 76CC | 盌 | ≡ | r-neither |  | identity |
| 76CC | 盌 | 7897 | 碗 | **→** | both |  |  |
| **←** | blocked |  |  |
| 76CC | 盌 | 92FA | 鋺 | ↔ | blocked |  |  |
| 7897 | 碗 | 7897 | 碗 | ≡ | r-both |  | identity |
| 7897 | 碗 | 92FA | 鋺 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 92FA | 鋺 | 92FA | 鋺 | ≡ | r-trad |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 6900 | 椀 | 6900 | 椀 | ≡ | r-both |  | identity |
| 6900 | 椀 | 76CC | 盌 | ↔ | blocked |  |  |
| 6900 | 椀 | 7897 | 碗 | ↔ | blocked |  |  |
| 6900 | 椀 | 92FA | 鋺 | ↔ | blocked |  |  |
| 76CC | 盌 | 76CC | 盌 | ≡ | r-both |  | identity |
| 76CC | 盌 | 7897 | 碗 | ↔ | blocked |  |  |
| 76CC | 盌 | 92FA | 鋺 | ↔ | blocked |  |  |
| 7897 | 碗 | 7897 | 碗 | ≡ | r-both |  | identity |
| 7897 | 碗 | 92FA | 鋺 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 92FA | 鋺 | 92FA | 鋺 | ≡ | r-trad |  | identity |

While in Unihan there is a kSemanticVariant mapping between U+6900 and U+7897, the ‘trad’ mapping suggested by CLGR13 seems dubious, and the dotAsia mapping seems preferable.

**Conclusion**: After further review by our CJK expert, the CLG13 set is acceptable.

The code point U+6A53 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (table) and dotAsia. CDNC 2018 has the same mapping as CLGR13. In dotAsia In those LGRs the two code points: U+6A53 and U+8563 are both singleton reflexive variant of type ‘r-both’. It should be noted that CLGR12 did not contain that variant set.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 6A53 | 橓 | 6A53 | 橓 | ≡ | r-neither |  | identity |
| 6A53 | 橓 | 8563 | 蕣 | **→** | both |  |  |
| **←** | blocked |
| 8563 | 蕣 | 8563 | 蕣 | ≡ | r-both |  | identity |

Unihan has a KSemanticVariant relationship between these two code points which would support the CLGR13 variant set existence.

**Conclusion**: After further review, the CLGR13 set is acceptable.

The code point U+6EDD was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 6CF7 | 泷 | 6CF7 | 泷 | ≡ | r-simp |  | identity |
| 6CF7 | 泷 | 6EDD | 滝 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 6CF7 | 泷 | 7027 | 瀧 | **→** | trad |  |  |
| **←** | simp |  |  |
| 6EDD | 滝 | 6EDD | 滝 | ≡ | r-neither |  | identity |
| 6EDD | 滝 | 7027 | 瀧 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 7027 | 瀧 | 7027 | 瀧 | ≡ | r-trad |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 6CF7 | 泷 | 6CF7 | 泷 | ≡ | r-simp |  | identity |
| 6CF7 | 泷 | 6EDD | 滝 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 6CF7 | 泷 | 7027 | 瀧 | **→** | trad |  |  |
| **←** | simp |  |  |
| 6EDD | 滝 | 6EDD | 滝 | ≡ | r-trad |  | identity |
| 6EDD | 滝 | 7027 | 瀧 | ↔ | blocked |  |  |
| 7027 | 瀧 | 7027 | 瀧 | ≡ | r-trad |  | identity |

There is no support in Unihan for a traditional mapping between U+6EDD and U+7027 as suggested by CLGR13.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+7200 was included in CLGR13 because of its membership in IICORE HKSCS but is treated differently between dotAsia and CLGR13 (table follows). In dotAsia both code points are singleton reflexive variants of type ‘r-both’.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7200 | 爀 | 7200 | 爀 | ≡ | r-neither |  | identity |
| 7200 | 爀 | 8D6B | 赫 | **→** | both |  |  |
| **←** | blocked |  |  |
| 8D6B | 赫 | 8D6B | 赫 | ≡ | r-both |  | identity |

Unihan does not have a variant relationship between these code points. However, they have the same kCantonese data: ‘haak1’.

**Conclusion**: After further by our CJK expert, the CLGR13 set is acceptable.

The code point U+734F was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 734F | 獏 | 734F | 獏 | ≡ | r-neither |  | identity |
| 734F | 獏 | 8C98 | 貘 | **→** | both |  |  |
| **←** | blocked |  |  |
| 8C98 | 貘 | 8C98 | 貘 | ≡ | r-both |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 734F | 獏 | 734F | 獏 | ≡ | r-trad |  | identity |
| 734F | 獏 | 8C98 | 貘 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 8C98 | 貘 | 8C98 | 貘 | ≡ | r-both |  | identity |

Unihan has Zvariant relationship between these code points and U+8C8A.

**Conclusion**: Because CDNC-2018 has the same definition as CLGR13, and after further review, this is acceptable.

The code point U+73E1 was included in CLGR13 because of its membership in IICORE HKSCS but is treated differently between dotAsia and CLGR13 (table follows). In dotAsia both code points are singleton reflexive variants of type ‘r-both’.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 73E1 | 珡 | 73E1 | 珡 | ≡ | r-neither |  | identity |
| 73E1 | 珡 | 7434 | 琴 | **→** | both |  |  |
| **←** | blocked |  |  |
| 7434 | 琴 | 7434 | 琴 | ≡ | r-both |  | identity |

Unihan has no formal variant relationship between these code points but U+73E1 in its kDefinition indicates a variant situation with U+7434.

**Conclusion**: Based on the above, the CLGR13 set is acceptable.

The code point U+764E was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 75EB | 痫 | 75EB | 痫 | ≡ | r-simp |  | identity |
| 75EB | 痫 | 7647 | 癇 | **→** | trad |  |  |
| **←** | simp |  |  |
| 75EB | 痫 | 764E | 癎 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 7647 | 癇 | 7647 | 癇 | ≡ | r-trad |  | identity |
| 7647 | 癇 | 764E | 癎 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 764E | 癎 | 764E | 癎 | ≡ | r-neither |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 75EB | 痫 | 75EB | 痫 | ≡ | r-simp |  | identity |
| 75EB | 痫 | 7647 | 癇 | **→** | trad |  |  |
| **←** | simp |  |  |
| 75EB | 痫 | 764E | 癎 | ↔ | blocked |  |  |
| 7647 | 癇 | 7647 | 癇 | ≡ | r-trad |  | identity |
| 7647 | 癇 | 764E | 癎 | ↔ | blocked |  |  |
| 764E | 癎 | 764E | 癎 | ≡ | r-both |  | identity |

Unihan has a trad-simp mapping between U+75EB and U+7647 and they all share a kCantonese value of ‘haan4’.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+89A9 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7779 | 睹 | 7779 | 睹 | ≡ | r-both |  | identity |
| 7779 | 睹 | 89A9 | 覩 | **→** | blocked |  |  |
| **←** | both |  |  |
| 89A9 | 覩 | 89A9 | 覩 | ≡ | r-neither |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7779 | 睹 | 7779 | 睹 | ≡ | r-both |  | identity |
| 7779 | 睹 | 89A9 | 覩 | ↔ | blocked |  |  |
| 89A9 | 覩 | 89A9 | 覩 | ≡ | r-both |  | identity |

Unihan has a KSemanticVariant between U+7779 and U+89A9. But as with many others, this does not determine whether U+89A9 should be ‘pre-empted’.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+7C83 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 79D5 | 秕 | 79D5 | 秕 | ≡ | r-both |  | identity |
| 79D5 | 秕 | 7C83 | 粃 | **→** | blocked |  |  |
| **←** | both |  |  |
| 7C83 | 粃 | 7C83 | 粃 | ≡ | r-neither |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 79D5 | 秕 | 79D5 | 秕 | ≡ | r-both |  | identity |
| 79D5 | 秕 | 7C83 | 粃 | ↔ | blocked |  |  |
| 7C83 | 粃 | 7C83 | 粃 | ≡ | r-both |  | identity |

Unihan has a KSemanticVariant between U+79D5 and U+7C83. But as with many others, this does not determine whether U+7C83 should be ‘pre-empted’.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+7B6F was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7B6F | 筯 | 7B6F | 筯 | ≡ | r-neither |  | identity |
| 7B6F | 筯 | 7BB8 | 箸 | **→** | both |  |  |
| **←** | blocked |  |  |
| 7BB8 | 箸 | 7BB8 | 箸 | ≡ | r-both |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7B6F | 筯 | 7B6F | 筯 | ≡ | r-both |  | identity |
| 7B6F | 筯 | 7BB8 | 箸 | ↔ | blocked |  |  |
| 7BB8 | 箸 | 7BB8 | 箸 | ≡ | r-both |  | identity |

Unihan has a KSemanticVariant between U+7B6F and U+7BB8. But as with many others, this does not determine whether U+7B6F should be ‘pre-empted’.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The variant set {U+7BC4, U+8303} is treated differently between CLGR13 (1st table) and both dotAsia and CDNC2018 (2nd table).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7BC4 | 範 | 7BC4 | 範 | ≡ | r-trad |  | identity |
| 7BC4 | 範 | 8303 | 范 | **→** | simp |  |  |
| **←** | trad-1 |  |  |
| 8303 | 范 | 8303 | 范 | ≡ | trad-2 |  | identity |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7BC4 | 範 | 7BC4 | 範 | ≡ | r-trad |  | identity |
| 7BC4 | 範 | 8303 | 范 | **→** | simp |  |  |
| **←** | trad |  |  |
| 8303 | 范 | 8303 | 范 | ≡ | r-both |  | identity |

The CLGR 13 table has some format issue (‘trad-2’ is not a reflexive type), and ‘r-both-mt’ may have been the intended mapping (this was the case in CLGR12). However the use of ‘trad-1’ and ‘trad-2’ may have been a trial at having two ‘trad’ mappings. There are no solutions to this unless new mapping types are created to allow equal footing allocatable variants.

**Conclusion**: This needs further review.

The code point U+7DDC was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7DBF | 綿 | 7DBF | 綿 | ≡ | r-trad |  | identity |
| 7DBF | 綿 | 7DDC | 緜 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 7DBF | 綿 | 7EF5 | 绵 | **→** | simp |  |  |
| **←** | trad |  |  |
| 7DDC | 緜 | 7DDC | 緜 | ≡ | r-neither |  | identity |
| 7DDC | 緜 | 7EF5 | 绵 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 7EF5 | 绵 | 7EF5 | 绵 | ≡ | r-simp |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7DBF | 綿 | 7DBF | 綿 | ≡ | r-trad |  | identity |
| 7DBF | 綿 | 7DDC | 緜 | ↔ | blocked |  |  |
| 7DBF | 綿 | 7EF5 | 绵 | **→** | simp |  |  |
| **←** | trad |  |  |
| 7DDC | 緜 | 7DDC | 緜 | ≡ | r-both |  | identity |
| 7DDC | 緜 | 7EF5 | 绵 | ↔ | blocked |  |  |
| 7EF5 | 绵 | 7EF5 | 绵 | ≡ | r-simp |  | identity |

Unihan has a KSemanticVariant between U+7DBF and U+7DDC.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+994D was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 81B3 | 膳 | 81B3 | 膳 | ≡ | r-both |  | identity |
| 81B3 | 膳 | 994D | 饍 | **→** | blocked |  |  |
| **←** | both |  |  |
| 994D | 饍 | 994D | 饍 | ≡ | r-neither |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 81B3 | 膳 | 81B3 | 膳 | ≡ | r-both |  | identity |
| 81B3 | 膳 | 994D | 饍 | ↔ | blocked |  |  |
| 994D | 饍 | 994D | 饍 | ≡ | r-both |  | identity |

Unihan has a KSemanticVariant between U+81B3 and U+994D.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+84DA was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 84DA | 蓚 | 84DA | 蓚 | ≡ | r-neither |  | identity |
| 84DA | 蓚 | 84E8 | 蓨 | **→** | both |  |  |
| **←** | blocked |  |  |
| 84E8 | 蓨 | 84E8 | 蓨 | ≡ | r-both |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 84DA | 蓚 | 84DA | 蓚 | ≡ | r-both |  | identity |
| 84DA | 蓚 | 84E8 | 蓨 | ↔ | blocked |  |  |
| 84E8 | 蓨 | 84E8 | 蓨 | ≡ | r-both |  | identity |

Unihan does not have variant relationship for these code points, however they have common kCantonese (tiu4), and kMandarin (tiáo) values.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+8EE2 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 8EE2 | 転 | 8EE2 | 転 | ≡ | r-neither |  | identity |
| 8EE2 | 転 | 8F49 | 轉 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 8EE2 | 転 | 8F6C | 转 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 8F49 | 轉 | 8F49 | 轉 | ≡ | r-trad |  | identity |
| 8F49 | 轉 | 8F6C | 转 | **→** | simp |  |  |
| **←** | trad |  |  |
| 8F6C | 转 | 8F6C | 转 | ≡ | r-simp |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 8EE2 | 転 | 8EE2 | 転 | ≡ | r-trad |  | identity |
| 8EE2 | 転 | 8F49 | 轉 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 8EE2 | 転 | 8F6C | 转 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 8F49 | 轉 | 8F49 | 轉 | ≡ | r-trad |  | identity |
| 8F49 | 轉 | 8F6C | 转 | **→** | simp |  |  |
| **←** | trad |  |  |
| 8F6C | 转 | 8F6C | 转 | ≡ | r-simp |  | identity |

In Unihan U+8EE2 is a zVariant of U+8F49.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+8FBA was included in CLGR13 because of its membership in IICORE HKSCS but is treated differently between dotAsia (where it is a singleton reflexive variant of type ‘r-both’) and CLGR13 (table follows).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 8FB9 | 边 | 8FB9 | 边 | ≡ | r-simp |  | identity |
| 8FB9 | 边 | 8FBA | 辺 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 8FB9 | 边 | 9089 | 邉 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 8FB9 | 边 | 908A | 邊 | **→** | trad |  |  |
| **←** | simp |  |  |
| 8FBA | 辺 | 8FBA | 辺 | ≡ | r-neither |  | identity |
| 8FBA | 辺 | 9089 | 邉 | ↔ | blocked |  |  |
| 8FBA | 辺 | 908A | 邊 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 9089 | 邉 | 9089 | 邉 | ≡ | r-neither |  | identity |
| 9089 | 邉 | 908A | 邊 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 908A | 邊 | 908A | 邊 | ≡ | r-trad |  | identity |

In Unihan U+8FBA is a zVariant of U+908A, therefore CLGR13 seems to make more sense.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+99C5 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR7 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 99C5 | 駅 | 99C5 | 駅 | ≡ | r-neither |  | identity |
| 99C5 | 駅 | 9A5B | 驛 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 99C5 | 駅 | 9A7F | 驿 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 9A5B | 驛 | 9A5B | 驛 | ≡ | r-trad |  | identity |
| 9A5B | 驛 | 9A7F | 驿 | **→** | simp |  |  |
| **←** | trad |  |  |
| 9A7F | 驿 | 9A7F | 驿 | ≡ | r-simp |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 99C5 | 駅 | 99C5 | 駅 | ≡ | r-trad |  | identity |
| 99C5 | 駅 | 9A5B | 驛 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 99C5 | 駅 | 9A7F | 驿 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 9A5B | 驛 | 9A5B | 驛 | ≡ | r-trad |  | identity |
| 9A5B | 驛 | 9A7F | 驿 | **→** | simp |  |  |
| **←** | trad |  |  |
| 9A7F | 驿 | 9A7F | 驿 | ≡ | r-simp |  | identity |

In Unihan, U+99C5 is a zVariant of U+9A5B.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+9D44 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 9D1F | 鴟 | 9D1F | 鴟 | ≡ | r-trad |  | identity |
| 9D1F | 鴟 | 9D44 | 鵄 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 9D1F | 鴟 | 9E31 | 鸱 | **→** | simp |  |  |
| **←** | trad |  |  |
| 9D44 | 鵄 | 9D44 | 鵄 | ≡ | r-neither |  | identity |
| 9D44 | 鵄 | 9E31 | 鸱 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 9E31 | 鸱 | 9E31 | 鸱 | ≡ | r-simp |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 9D1F | 鴟 | 9D1F | 鴟 | ≡ | r-trad |  | identity |
| 9D1F | 鴟 | 9D44 | 鵄 | ↔ | blocked |  |  |
| 9D1F | 鴟 | 9E31 | 鸱 | **→** | simp |  |  |
| **←** | trad |  |  |
| 9D44 | 鵄 | 9D44 | 鵄 | ≡ | r-both |  | identity |
| 9D44 | 鵄 | 9E31 | 鸱 | ↔ | blocked |  |  |
| 9E31 | 鸱 | 9E31 | 鸱 | ≡ | r-simp |  | identity |

Unihan indicates that U+9D1F and U+9E31 are in a trad-simp relationship and that U+9D44 is a kDefinition variant of U+9D1F, but this does not provide a clear answer to determine which of the two LGRs is better.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+9F62 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 9F61 | 齡 | 9F61 | 齡 | ≡ | r-trad |  | identity |
| 9F61 | 齡 | 9F62 | 齢 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 9F61 | 齡 | 9F84 | 龄 | **→** | simp |  |  |
| **←** | trad |  |  |
| 9F62 | 齢 | 9F62 | 齢 | ≡ | r-neither |  | identity |
| 9F62 | 齢 | 9F84 | 龄 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 9F84 | 龄 | 9F84 | 龄 | ≡ | r-simp |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 9F61 | 齡 | 9F61 | 齡 | ≡ | r-trad |  | identity |
| 9F61 | 齡 | 9F62 | 齢 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 9F61 | 齡 | 9F84 | 龄 | **→** | simp |  |  |
| **←** | trad |  |  |
| 9F62 | 齢 | 9F62 | 齢 | ≡ | r-trad |  | identity |
| 9F62 | 齢 | 9F84 | 龄 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 9F84 | 龄 | 9F84 | 龄 | ≡ | r-simp |  | identity |

In Unihan, U+9FC2 is a zVariant of U+9F61.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

## Detailed list of differences for dotAsia additions – 19 sets

The following numbered items list all the information available to the IP at this point for each of the differences found and its recommendations based on evidence and the opinion of the IP CJK expert.

Case #1 (U+39DB㧛U+3A5C㩜U+63FD揽U+64E5擥U+652C攬) is a mix of cases involving three added code points: U+3A5C (added in CNDC 2018 but not in dotAsia), U+64E5 (added in both CDNC 2018 and dotAsia), and U+39DB (only added in dotAsia). This complex case is examined separately in 8.1.

The code point U+3A18 was included in CLGR13 because of its membership in IICORE HKSCS but is has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 3A18 | 㨘 | 3A18 | 㨘 | ≡ | r-trad |  | identity |
| 3A18 | 㨘 | 64E4 | 擤 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 64E4 | 擤 | 64E4 | 擤 | ≡ | r-both |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 3A18 | 㨘 | 3A18 | 㨘 | ≡ | r-both |  | identity |
| 3A18 | 㨘 | 64E4 | 擤 | ↔ | blocked |  |  |
| 64E4 | 擤 | 64E4 | 擤 | ≡ | r-both |  | identity |

Unihan in the entry for U+3A18 mentions in its kDefinition field that it is “(same as U+64E4 擤) to blow the nose with the fingers; (Cant.) to scour; to rebuke; to hit with a ball”; but it has no traditional/simplified mapping.

Therefore, while a semantic variant is implied, the traditional/simplified mapping added by CLGR13 is not supported.

**Conclusion**: The dotAsia definition is preferred by our CJK expert.

The code point U+3A52 was included in CLGR13 because of its membership in IICORE HKSCS but is treated differently between CLGR13 (table follows) and dotAsia (where it is a singleton reflexive variant of ‘r-both’). CDNC-2018 does not include U+3A52.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 3A52 | 㩒 | 3A52 | 㩒 | ≡ | r-trad |  | identity |
| 3A52 | 㩒 | 64D2 | 擒 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 64D2 | 擒 | 64D2 | 擒 | ≡ | r-both |  | identity |

The simplified mapping between U+3A52 and U+64D2 in CLGR13 is not supported by Unihan and looks doubtful. Unihan kDefinition field for U+3A52 indicates that this is a variant of U+64D2 but without simplified mapping. In Unihan U+64D2 has itself a semantic variant relationship with U+6366 捦, not supported by either CLGR13 or dotAsia.

**Conclusion**: After further review, our CJK expert found the CLGR13 set acceptable.

The code point U+4C7D was included in CLGR13 because of its membership in IICORE HKSCS and the code point U+4C9D was added because of its GS reference (Singapore source). In the CLGR13 (1st), U+4C7D and U+4C9D are part of a larger variant set. In dotAsia (2nd), U+4C7D and U+4C9D form a smaller variant set together with compatible values; similarly, the pair U+9BE7 and U+9CB3 forms another variant set with compatible values. See 4.3.3.2 for further details.

The code point U+701E was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between dotAsia (table follows) where it is part of the variant set U+51C0, U+51C8, U+6D44, U+6DE8, and U+701E. In CLGR13, U+701E is a singleton reflexive variant of type ‘r-both.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 51C0 | 净 | 51C0 | 净 | ≡ | r-simp |  | identity |
| 51C0 | 净 | 51C8 | 凈 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 51C0 | 净 | 6D44 | 浄 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 51C0 | 净 | 6DE8 | 淨 | **→** | trad |  |  |
| **←** | simp |  |  |
| 51C0 | 净 | 701E | 瀞 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 51C8 | 凈 | 51C8 | 凈 | ≡ | r-trad |  | identity |
| 51C8 | 凈 | 6D44 | 浄 | ↔ | blocked |  |  |
| 51C8 | 凈 | 6DE8 | 淨 | ↔ | blocked |  |  |
| 51C8 | 凈 | 701E | 瀞 | ↔ | blocked |  |  |
| 6D44 | 浄 | 6D44 | 浄 | ≡ | r-neither |  | identity |
| 6D44 | 浄 | 6DE8 | 淨 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 6D44 | 浄 | 701E | 瀞 | ↔ | blocked |  |  |
| 6DE8 | 淨 | 6DE8 | 淨 | ≡ | r-trad |  | identity |
| 6DE8 | 淨 | 701E | 瀞 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 701E | 瀞 | 701E | 瀞 | ≡ | r-neither |  | identity |

Unihan has variant relationship between U+51C0, U+51C8, U+6D44, and U+6DE8, but not U+701E.

**Conclusion**: Our CJK expert prefers the dotAsia definition.

The code point U+6335 was included in CLGR13 because of its membership in IICORE HKSCS set but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5F04 | 弄 | 5F04 | 弄 | ≡ | r-both |  | identity |
| 5F04 | 弄 | 6335 | 挵 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 5F04 | 弄 | 8856 | 衖 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 6335 | 挵 | 6335 | 挵 | ≡ | r-trad |  | identity |
| 6335 | 挵 | 8856 | 衖 | ↔ | blocked |  |  |
| 8856 | 衖 | 8856 | 衖 | ≡ | r-trad |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 5F04 | 弄 | 5F04 | 弄 | ≡ | r-both |  | identity |
| 5F04 | 弄 | 6335 | 挵 | ↔ | blocked |  |  |
| 5F04 | 弄 | 8856 | 衖 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 6335 | 挵 | 6335 | 挵 | ≡ | r-both |  | identity |
| 6335 | 挵 | 8856 | 衖 | ↔ | blocked |  |  |
| 8856 | 衖 | 8856 | 衖 | ≡ | r-trad |  | identity |

Unihan does not bring any clarification (U+5F04 has variant relationship with U+7627 and U+759F which is a separate CLGR13 variant set, but U+6335 and U+8856 have no variant in Unihan).

**Conclusion**: Our CJK expert prefers the CLGR13 definition.

The code point U+60E3 was included in CLGR13 because of its membership in IICORE HKSCS; however, it is treated differently. In dotAsia, it is part of another variant set with U+63D4 (2nd table). In CLGR13, it is added to the set U+603B, U+6374, U+6403, U+6460, U+7DCF, U+7E02 and U+7E3D (1st table), and U+63D4 is a singleton reflexive variant reflexive of type ‘both’; all other mappings identical to dotAsia.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 603B | 总 | 603B | 总 | ≡ | r-simp |  | identity |
| 603B | 总 | 60E3 | 惣 | **→** | blocked |  |  |
| **←** | simp |
| 603B | 总 | 6374 | 捴 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 603B | 总 | 6403 | 搃 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 603B | 总 | 6460 | 摠 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 603B | 总 | 7DCF | 総 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 603B | 总 | 7E02 | 縂 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 603B | 总 | 7E3D | 總 | **→** | trad |  |  |
| **←** | simp |  |  |
| 60E3 | 惣 | 60E3 | 惣 | ≡ | r-neither |  | identity |
| 60E3 | 惣 | 6374 | 捴 | ↔ | blocked |  |  |
| 60E3 | 惣 | 6403 | 搃 | ↔ | blocked |  |  |
| 60E3 | 惣 | 6460 | 摠 | ↔ | blocked |  |  |
| 60E3 | 惣 | 7DCF | 総 | ↔ | blocked |  |  |
| 60E3 | 惣 | 7E02 | 縂 | ↔ | blocked |  |  |
| 60E3 | 惣 | 7E3D | 總 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 6374 | 捴 | 6374 | 捴 | ≡ | r-neither |  | identity |
| 6374 | 捴 | 6403 | 搃 | ↔ | blocked |  |  |
| 6374 | 捴 | 6460 | 摠 | ↔ | blocked |  |  |
| 6374 | 捴 | 7DCF | 総 | ↔ | blocked |  |  |
| 6374 | 捴 | 7E02 | 縂 | ↔ | blocked |  |  |
| 6374 | 捴 | 7E3D | 總 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 6403 | 搃 | 6403 | 搃 | ≡ | r-neither |  | identity |
| 6403 | 搃 | 6460 | 摠 | ↔ | blocked |  |  |
| 6403 | 搃 | 7DCF | 総 | ↔ | blocked |  |  |
| 6403 | 搃 | 7E02 | 縂 | ↔ | blocked |  |  |
| 6403 | 搃 | 7E3D | 總 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 6460 | 摠 | 6460 | 摠 | ≡ | r-trad |  | identity |
| 6460 | 摠 | 7DCF | 総 | ↔ | blocked |  |  |
| 6460 | 摠 | 7E02 | 縂 | ↔ | blocked |  |  |
| 6460 | 摠 | 7E3D | 總 | ↔ | blocked |  |  |
| 7DCF | 総 | 7DCF | 総 | ≡ | r-neither |  | identity |
| 7DCF | 総 | 7E02 | 縂 | ↔ | blocked |  |  |
| 7DCF | 総 | 7E3D | 總 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 7E02 | 縂 | 7E02 | 縂 | ≡ | r-neither |  | identity |
| 7E02 | 縂 | 7E3D | 總 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 7E3D | 總 | 7E3D | 總 | ≡ | r-trad |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 60E3 | 惣 | 60E3 | 惣 | ≡ | r-trad |  | identity |
| 60E3 | 惣 | 63D4 | 揔 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 63D4 | 揔 | 63D4 | 揔 | ≡ | r-both |  | identity |

Unihan has a variant relationship between U+603B, U+6374, U+6403, U+6460, U+7DCF, U+7E02, and U+7E3D (U+603B and U+7E3D are trad-simp), but has no variant mapping for U+60E3 or U+63D4 (although they share the same reading in Mandarin).

**Conclusion**: After further review, our CJK expert found the CLGR13 set acceptable.

The code point U+656D was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd). Note that U+656D is mostly used in a Korean context.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 626C | 扬 | 626C | 扬 | ≡ | r-simp |  | identity |
| 626C | 扬 | 63DA | 揚 | **→** | trad |  |  |
| **←** | simp |  |  |
| 626C | 扬 | 656D | 敭 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 626C | 扬 | 98BA | 颺 | ↔ | blocked |  |  |
| 626C | 扬 | 98CF | 飏 | ↔ | blocked |  |  |
| 63DA | 揚 | 63DA | 揚 | ≡ | r-trad |  | identity |
| 63DA | 揚 | 656D | 敭 | ↔ | blocked |  |  |
| 63DA | 揚 | 98BA | 颺 | ↔ | blocked |  |  |
| 63DA | 揚 | 98CF | 飏 | ↔ | blocked |  |  |
| 656D | 敭 | 656D | 敭 | ≡ | r-trad |  | identity |
| 656D | 敭 | 98BA | 颺 | ↔ | blocked |  |  |
| 656D | 敭 | 98CF | 飏 | ↔ | blocked |  |  |
| 98BA | 颺 | 98BA | 颺 | ≡ | r-trad |  | identity |
| 98BA | 颺 | 98CF | 飏 | **→** | simp |  |  |
| **←** | trad |  |  |
| 98CF | 飏 | 98CF | 飏 | ≡ | r-simp |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 626C | 扬 | 626C | 扬 | ≡ | r-simp |  | identity |
| 626C | 扬 | 63DA | 揚 | **→** | trad |  |  |
| **←** | simp |  |  |
| 626C | 扬 | 656D | 敭 | ↔ | blocked |  |  |
| 626C | 扬 | 98BA | 颺 | ↔ | blocked |  |  |
| 626C | 扬 | 98CF | 飏 | ↔ | blocked |  |  |
| 63DA | 揚 | 63DA | 揚 | ≡ | r-trad |  | identity |
| 63DA | 揚 | 656D | 敭 | ↔ | blocked |  |  |
| 63DA | 揚 | 98BA | 颺 | ↔ | blocked |  |  |
| 63DA | 揚 | 98CF | 飏 | ↔ | blocked |  |  |
| 656D | 敭 | 656D | 敭 | ≡ | r-both |  | identity |
| 656D | 敭 | 98BA | 颺 | ↔ | blocked |  |  |
| 656D | 敭 | 98CF | 飏 | ↔ | blocked |  |  |
| 98BA | 颺 | 98BA | 颺 | ≡ | r-trad |  | identity |
| 98BA | 颺 | 98CF | 飏 | **→** | simp |  |  |
| **←** | trad |  |  |
| 98CF | 飏 | 98CF | 飏 | ≡ | r-simp |  | identity |

Unihan has simp-trad variant relationship between U+626C and U+63DA, and another relationship between U+98BA and U+98CF, and none for U+656D.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+637F was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between dotAsia and CLGR13. In CLGR13 (1st table) it is part of a variant relationship with U+6816 and U+68F2. In dotAsia, U+637F is a singleton reflexive variant of type ‘r-both’ and the two other code points: U+6816 and U+68F2 (2nd table) are in a variant relationship.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Source | | Glyph | | Target | | Glyph | |  | | Type(s) | | Ref | | Comment | |
| 637F | | 捿 | | 637F | | 捿 | | ≡ | | r-neither | |  | | identity | |
| 637F | | 捿 | | 6816 | | 栖 | | **→** | | simp | |  | |  | |
| **←** | | blocked | |  | |  | |
| 637F | | 捿 | | 68F2 | | 棲 | | **→** | | trad | |  | |  | |
| **←** | | blocked | |
| 6816 | | 栖 | | 6816 | | 栖 | | ≡ | | r-both | |  | |  | |
| 6816 | | 栖 | | 68F2 | | 棲 | | **→** | | trad | |  | |  | |
| **←** | | blocked | |
| 68F2 | | 棲 | | 68F2 | | 棲 | | ≡ | | r-both | |  | | identity | |
| **Source** | | **Glyph** | | **Target** | | **Glyph** | |  | | **Type(s)** | | **Ref** | | **Comment** | |
| 6816 | | 栖 | | 6816 | | 栖 | | ≡ | | r-both | |  | | identity | |
| 6816 | | 栖 | | 68F2 | | 棲 | | **→** | | trad | |  | |  | |
| **←** | | simp | |  | |  | |
| 68F2 | | 棲 | | 68F2 | | 棲 | | ≡ | | r-trad | |  | | identity | |

Unihan has a trad-simp variant relationship between U+6816 and U+68F2, but nothing for 637F. This would support the dotAsia mappings. U+637F uses a different radical that is just visual similar, but easily distinguishable by most CJK users.

CDNC-2018 has the same variant set as dotAsia for U+6816 and U+68F2 (2nd table) and does not include U+637F. However there is an issue with the dotAsia because it contains mapping values from U+6816 of ‘r-both’ and ‘trad’ which cannot be maintained as it could create a large number of variants.

The simplest solution is to transform ‘r-both’ into ‘r-both-mt’ which is equivalent to ‘r-simp’, but that makes U+6816 the ‘preferred’ traditional character which is not necessarily desired.

**Conclusion**: After further review by our CJK expert, the CLGR13 was found to be acceptable.

The code point U+6667 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st table) and dotAsia (2nd table).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 6667 | 晧 | 6667 | 晧 | ≡ | r-neither |  | identity |
| 6667 | 晧 | 66A0 | 暠 | ↔ | blocked |  |  |
| 6667 | 晧 | 7693 | 皓 | **→** | both |  |  |
| **←** | blocked |
| 6667 | 晧 | 769C | 皜 | ↔ | blocked |  |  |
| 66A0 | 暠 | 66A0 | 暠 | ≡ | r-trad |  | identity |
| 66A0 | 暠 | 7693 | 皓 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 66A0 | 暠 | 769C | 皜 | ↔ | blocked |  |  |
| 7693 | 皓 | 7693 | 皓 | ≡ | r-both |  | identity |
| 7693 | 皓 | 769C | 皜 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 769C | 皜 | 769C | 皜 | ≡ | r-trad |  | identity |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 6667 | 晧 | 6667 | 晧 | ≡ | r-both |  | identity |
| 6667 | 晧 | 66A0 | 暠 | ↔ | blocked |  |  |
| 6667 | 晧 | 7693 | 皓 | ↔ | blocked |  |  |
| 6667 | 晧 | 769C | 皜 | ↔ | blocked |  |  |
| 66A0 | 暠 | 66A0 | 暠 | ≡ | r-trad |  | identity |
| 66A0 | 暠 | 7693 | 皓 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 66A0 | 暠 | 769C | 皜 | ↔ | blocked |  |  |
| 7693 | 皓 | 7693 | 皓 | ≡ | r-both |  | identity |
| 7693 | 皓 | 769C | 皜 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 769C | 皜 | 769C | 皜 | ≡ | r-trad |  | identity |

Unihan has a variant relationship between U+66A0 and U+769C. It has none for U+6667, and the last element U+7693 has mapping with other code points (U+769D, U+9865, and U+98A2).

**Conclusion**: After further review, the CLGR13 set is acceptable.

The code point U+7AC3 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7076 | 灶 | 7076 | 灶 | ≡ | r-both |  | identity |
| 7076 | 灶 | 7AC3 | 竃 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 7076 | 灶 | 7AC8 | 竈 | **→** | blocked |  |  |
| **←** | both |  |  |
| 7AC3 | 竃 | 7AC3 | 竃 | ≡ | r-neither |  | identity |
| 7AC3 | 竃 | 7AC8 | 竈 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 7AC8 | 竈 | 7AC8 | 竈 | ≡ | r-neither |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7076 | 灶 | 7076 | 灶 | ≡ | r-both |  | identity |
| 7076 | 灶 | 7AC3 | 竃 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 7076 | 灶 | 7AC8 | 竈 | **→** | blocked |  |  |
| **←** | both |  |  |
| 7AC3 | 竃 | 7AC3 | 竃 | ≡ | r-trad |  | identity |
| 7AC3 | 竃 | 7AC8 | 竈 | ↔ | blocked |  |  |
| 7AC8 | 竈 | 7AC8 | 竈 | ≡ | r-neither |  | identity |

Unihan has a variant relationship between all code points.

**Conclusion**: After further review, the CLGR13 is acceptable.

The code point U+932C was included in CLGR13 because of its membership in IICORE HKSCS but is treated differently between dotAsia and CLGR13. In CLGR13, it is part of a large variant set (1st table). In dotAsia, it is a singleton reflexive of type ‘both’, while the members related to the CLGR13 variant set are treated differently (2nd table). The CDNC 2018 variant set is the same as dotAsia. U+932C is not part of CDNC 2018.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 70BC | 炼 | 70BC | 炼 | ≡ | r-simp |  | identity |
| 70BC | 炼 | 7149 | 煉 | **→** | trad |  |  |
| **←** | simp |  |  |
| 70BC | 炼 | 932C | 錬 | **→** | blocked |  |  |
| **←** | simp |
| 70BC | 炼 | 934A | 鍊 | ↔ | blocked |  |  |
| 7149 | 煉 | 7149 | 煉 | ≡ | r-trad |  | identity |
| 7149 | 煉 | 932C | 錬 | **→** | blocked |  |  |
| **←** | trad |
| 7149 | 煉 | 934A | 鍊 | ↔ | blocked |  |  |
| 932C | 錬 | 932C | 錬 | ≡ | r-neither |  | identity |
| 932C | 錬 | 934A | 鍊 | ↔ | blocked |  |  |
| 934A | 鍊 | 934A | 鍊 | ≡ | r-both |  | identity |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 70BC | 炼 | 70BC | 炼 | ≡ | r-simp |  | identity |
| 70BC | 炼 | 7149 | 煉 | **→** | trad |  |  |
| **←** | simp |  |  |
| 70BC | 炼 | 934A | 鍊 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 7149 | 煉 | 7149 | 煉 | ≡ | r-trad |  | identity |
| 7149 | 煉 | 934A | 鍊 | ↔ | blocked |  |  |
| 934A | 鍊 | 934A | 鍊 | ≡ | r-trad |  | identity |

Unihan has a trad-simp variant relationship between U+70BC and U+7149, and a zVariant with U+932C, but not with U+934A. According to our CJK expert, the rationale for adding U+932C to the variant set is weak; therefore it does not need to be added.

**Conclusion**: After further review, our CJK expert found the CLGR13 set acceptable.

The code point U+7460 was included in CLGR13 because of its membership in IICORE HKSCS but is treated differently between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7409 | 琉 | 7409 | 琉 | ≡ | r-both |  | identity |
| 7409 | 琉 | 7460 | 瑠 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 7460 | 瑠 | 7460 | 瑠 | ≡ | r-trad |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 7409 | 琉 | 7409 | 琉 | ≡ | r-both |  | identity |
| 7409 | 琉 | 7460 | 瑠 | ↔ | blocked |  |  |
| 7460 | 瑠 | 7460 | 瑠 | ≡ | r-both |  | identity |

The correlation between U+7409 and U+7460 is confirmed by Unihan (kSemanticVariant), but the difference in mappings cannot be clarified.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+74C8 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 740D | 琍 | 740D | 琍 | ≡ | r-trad |  | identity |
| 740D | 琍 | 7483 | 璃 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 740D | 琍 | 74C8 | 瓈 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 7483 | 璃 | 7483 | 璃 | ≡ | r-both |  | identity |
| 7483 | 璃 | 74C8 | 瓈 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 74C8 | 瓈 | 74C8 | 瓈 | ≡ | r-neither |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 740D | 琍 | 740D | 琍 | ≡ | r-trad |  | identity |
| 740D | 琍 | 7483 | 璃 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 740D | 琍 | 74C8 | 瓈 | ↔ | blocked |  |  |
| 7483 | 璃 | 7483 | 璃 | ≡ | r-both |  | identity |
| 7483 | 璃 | 74C8 | 瓈 | ↔ | blocked |  |  |
| 74C8 | 瓈 | 74C8 | 瓈 | ≡ | r-both |  | identity |

Although Unihan describes a kSemanticVariant between U+74C8 and U+7483, and U+740D shares the kCantonese value of ‘lei4’ with the two other code points. However, there is no traditional or simplified mapping implied for those code points.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+757A was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 757A | 畺 | 757A | 畺 | ≡ | r-neither |  | identity |
| 757A | 畺 | 7586 | 疆 | **→** | both |  |  |
| **←** | blocked |  |  |
| 7586 | 疆 | 7586 | 疆 | ≡ | r-both |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 757A | 畺 | 757A | 畺 | ≡ | r-both |  | identity |
| 757A | 畺 | 7586 | 疆 | ↔ | blocked |  |  |
| 7586 | 疆 | 7586 | 疆 | ≡ | r-both |  | identity |

Unihan has a KSemanticVariant between U+757A and U+7586.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+9771 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 8095 | 肕 | 8095 | 肕 | ≡ | r-trad |  | identity |
| 8095 | 肕 | 976D | 靭 | ↔ | blocked |  |  |
| 8095 | 肕 | 9771 | 靱 | ↔ | blocked |  |  |
| 8095 | 肕 | 97CC | 韌 | ↔ | blocked |  |  |
| 8095 | 肕 | 97E7 | 韧 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 976D | 靭 | 976D | 靭 | ≡ | r-neither |  | identity |
| 976D | 靭 | 9771 | 靱 | ↔ | blocked |  |  |
| 976D | 靭 | 97CC | 韌 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 976D | 靭 | 97E7 | 韧 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 9771 | 靱 | 9771 | 靱 | ≡ | r-neither |  | identity |
| 9771 | 靱 | 97CC | 韌 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 9771 | 靱 | 97E7 | 韧 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 97CC | 韌 | 97CC | 韌 | ≡ | r-trad |  | identity |
| 97CC | 韌 | 97E7 | 韧 | **→** | simp |  |  |
| **←** | trad |  |  |
| 97E7 | 韧 | 97E7 | 韧 | ≡ | r-simp |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 8095 | 肕 | 8095 | 肕 | ≡ | r-trad |  | identity |
| 8095 | 肕 | 976D | 靭 | ↔ | blocked |  |  |
| 8095 | 肕 | 9771 | 靱 | ↔ | blocked |  |  |
| 8095 | 肕 | 97CC | 韌 | ↔ | blocked |  |  |
| 8095 | 肕 | 97E7 | 韧 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 976D | 靭 | 976D | 靭 | ≡ | r-neither |  | identity |
| 976D | 靭 | 9771 | 靱 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 976D | 靭 | 97CC | 韌 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 976D | 靭 | 97E7 | 韧 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 9771 | 靱 | 9771 | 靱 | ≡ | r-neither |  | identity |
| 9771 | 靱 | 97CC | 韌 | ↔ | blocked |  |  |
| 9771 | 靱 | 97E7 | 韧 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 97CC | 韌 | 97CC | 韌 | ≡ | r-trad |  | identity |
| 97CC | 韌 | 97E7 | 韧 | **→** | simp |  |  |
| **←** | trad |  |  |
| 97E7 | 韧 | 97E7 | 韧 | ≡ | r-simp |  | identity |

In Unihan, U+9771 is a zVariant of U+97CC. The variation in ‘trad’ mapping between the two LGRs for U+9771 cannot be determined.

The original dotAsia table had no symmetry mapping from U+976D to U+9771 which was fixed by adding a ‘blocked’ mapping in this table comparison.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+8420 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between dotAsia (table follows) and CLGR13 (where it is a singleton reflexive variant of ‘r-both’).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 840C | 萌 | 840C | 萌 | ≡ | r-both |  | identity |
| 840C | 萌 | 8420 | 萠 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 840C | 萌 | 8544 | 蕄 | **→** | blocked |  |  |
| **←** | simp |  |  |
| 8420 | 萠 | 8420 | 萠 | ≡ | r-trad |  | identity |
| 8420 | 萠 | 8544 | 蕄 | ↔ | blocked |  |  |
| 8544 | 蕄 | 8544 | 蕄 | ≡ | r-trad |  | identity |

In Unihan U+8420 is a zVariant of U+U+840C, therefore some variant mapping is expected (as done in the dotAsia table). The dotAsia table seems preferable.

**Conclusion**: Our CJK expert prefers the dotAsia definition.

The code point U+9244 from was included in the CLGR13 (1st) because of its membership in IICORE HKSCS and is mapped differently from dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 9244 | 鉄 | 9244 | 鉄 | ≡ | r-neither |  | identity |
| 9244 | 鉄 | 9295 | 銕 | ↔ | blocked |  |  |
| 9244 | 鉄 | 9435 | 鐵 | **→** | trad |  |  |
| **←** | blocked |  |  |
| 9244 | 鉄 | 94C1 | 铁 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 9295 | 銕 | 9295 | 銕 | ≡ | r-trad |  | identity |
| 9295 | 銕 | 9435 | 鐵 | ↔ | blocked |  |  |
| 9295 | 銕 | 94C1 | 铁 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 9435 | 鐵 | 9435 | 鐵 | ≡ | r-trad |  | identity |
| 9435 | 鐵 | 94C1 | 铁 | **→** | simp |  |  |
| **←** | trad |  |  |
| 94C1 | 铁 | 94C1 | 铁 | ≡ | r-simp |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 9244 | 鉄 | 9244 | 鉄 | ≡ | r-trad |  | identity |
| 9244 | 鉄 | 9295 | 銕 | ↔ | blocked |  |  |
| 9244 | 鉄 | 9435 | 鐵 | ↔ | blocked |  |  |
| 9244 | 鉄 | 94C1 | 铁 | **→** | simp |  |  |
| **←** | trad |  |  |
| 9295 | 銕 | 9295 | 銕 | ≡ | r-trad |  | identity |
| 9295 | 銕 | 9435 | 鐵 | ↔ | blocked |  |  |
| 9295 | 銕 | 94C1 | 铁 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 9435 | 鐵 | 9435 | 鐵 | ≡ | r-trad |  | identity |
| 9435 | 鐵 | 94C1 | 铁 | **→** | simp |  |  |
| **←** | trad |  |  |
| 94C1 | 铁 | 94C1 | 铁 | ≡ | r-simp |  | identity |

The differences concerning U+9244 mappings are not addressed at this point (Unihan has Semantic Variant mappings between U+9244, U+9295, and U+9435).

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

The code point U+98C7 was included in CLGR13 because of its membership in IICORE HKSCS but has been assigned different types for its variant mappings between CLGR13 (1st) and dotAsia (2nd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 98C6 | 飆 | 98C6 | 飆 | ≡ | r-trad |  | identity |
| 98C6 | 飆 | 98C7 | 飇 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 98C6 | 飆 | 98C8 | 飈 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 98C6 | 飆 | 98D9 | 飙 | **→** | simp |  |  |
| **←** | trad |  |  |
| 98C6 | 飆 | 98DA | 飚 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 98C7 | 飇 | 98C7 | 飇 | ≡ | r-neither |  | identity |
| 98C7 | 飇 | 98C8 | 飈 | ↔ | blocked |  |  |
| 98C7 | 飇 | 98D9 | 飙 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 98C7 | 飇 | 98DA | 飚 | ↔ | blocked |  |  |
| 98C8 | 飈 | 98C8 | 飈 | ≡ | r-neither |  | identity |
| 98C8 | 飈 | 98D9 | 飙 | ↔ | blocked |  |  |
| 98C8 | 飈 | 98DA | 飚 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 98D9 | 飙 | 98D9 | 飙 | ≡ | r-simp |  | identity |
| 98D9 | 飙 | 98DA | 飚 | ↔ | blocked |  |  |
| 98DA | 飚 | 98DA | 飚 | ≡ | r-simp |  | identity |
| **Source** | **Glyph** | **Target** | **Glyph** |  | **Type(s)** | **Ref** | **Comment** |
| 98C6 | 飆 | 98C6 | 飆 | ≡ | r-trad |  | identity |
| 98C6 | 飆 | 98C7 | 飇 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 98C6 | 飆 | 98C8 | 飈 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 98C6 | 飆 | 98D9 | 飙 | **→** | simp |  |  |
| **←** | trad |  |  |
| 98C6 | 飆 | 98DA | 飚 | **→** | blocked |  |  |
| **←** | trad |  |  |
| 98C7 | 飇 | 98C7 | 飇 | ≡ | r-neither |  | identity |
| 98C7 | 飇 | 98C8 | 飈 | ↔ | blocked |  |  |
| 98C7 | 飇 | 98D9 | 飙 | ↔ | blocked |  |  |
| 98C7 | 飇 | 98DA | 飚 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 98C8 | 飈 | 98C8 | 飈 | ≡ | r-neither |  | identity |
| 98C8 | 飈 | 98D9 | 飙 | ↔ | blocked |  |  |
| 98C8 | 飈 | 98DA | 飚 | **→** | simp |  |  |
| **←** | blocked |  |  |
| 98D9 | 飙 | 98D9 | 飙 | ≡ | r-simp |  | identity |
| 98D9 | 飙 | 98DA | 飚 | ↔ | blocked |  |  |
| 98DA | 飚 | 98DA | 飚 | ≡ | r-simp |  | identity |

Unihan has two trad-simp variant relationship between the pairs (U+98C6, U+98D9) and (U+98C8, U+98DA), but U+98C7 is related to other code points.

**Conclusion**: After further review by our CJK expert, the CLGR13 set is acceptable.

# Background concerning visual exchangeability

CGP has been faced recently with the issue of visual similarity among CJK ideographs. This has been clarified as a visual exchangeability through additional discussions between CGP and IP. By contrast, “visual similarity” is a much broader issue, which is not addressed by data set in a LGR, but by some process beyond the LGR system. Visual exchangeability means that the glyphs in question create a security risk by not being distinguishable by the user community using these glyphs in the vast majority of usage scenarios.

It should also be noted that the process of CJK Ideographs unification is tightly connected to this concern. For example, CJK Unification as a principle allows ideographs from some sources to be separately encoded even if they are unifiable for reasons such as being visually confusable and cognate (term loosely used in CJK Unification context to represent semantic equivalence).

ISO/IEC 10646 5th edition (available at <http://standards.iso.org/ittf/PubliclyAvailableStandards/c069119_ISO_IEC_10646_2017.zip> ) contains in its Annex S lists of source separation examples (S.3) and non-unification examples (S.4). Typically, source separated ideographs are part of the same variant set while non-cognate characters that would otherwise be visually identical are not (but there are exception).

The Unicode Consortium has published a list of confusable characters (latest as <https://www.unicode.org/Public/security/11.0.0/confusables.txt> , including CJK ideographs which intersect some of the principles above without being a perfect match for any of them.

In the [Proposal], section 7, CGP includes these terms:

Due to nature of the Chinese writing system, a set of Chinese variant characters generally share the same radical or components, and thus have a certain degree of visual similarity, allowing educated readers to recognize the variant relationship easily. The semantically identical Chinese variants generally have visually similar forms (like 4443䑃 and 6726朦), but the reverse is not true, some visually similar Chinese characters have totally different semantics (like 58AB[墫] 58FF[壿]), typically, the Chinese script users don’t treat these characters as exchangeable variants.

However, it is quite common that semantically identical Chinese variants have very different visual forms. It is also not a requirement for any given variant set to have characters that are semantically identical, for example in the variant set {U+670F 朏, U+80D0 胐}, the 2 characters are not expected to be cognate (ref 10646 Annex S.4). So in effect, the current variant sets are already a mix of semantically identical characters and visually exchangeable characters (carefully avoiding the term ‘visual similarity’). There is no dispute about the fact that U+58AB 墫 and U+58FF 壿are not cognate, but there may be a debate about whether these 2 characters are exchangeable. The only difference consists in the radical on the left being either the earth or scholar which are barely distinguishable in that context.

Finally, the IP notes the following statement in the [Proposal]:

*But given CJK three parties haven’t reached consensus on the issue of “visually identical characters”, especially, the cross-script visually identical characters (Kana-Kanji, Hangul-Hanja), to handle the issue of visual similarity of strings causing confusability between (potential) TLDs, CJK GPs would like to see the issue be resolved by String Similarity Panels during TLD string evaluation, String Confusion Objection and String Contention Mechanisms or by user application tools such as browsers, as expressed in the public letter at <https://www.icann.org/en/system/files/correspondence/wei-et-al-to-marby-25jan19-en.pdf>. Hence, in this version of proposal document, CGP didn’t create the visual identical tag and corresponding rule for it.*

That letter was answered by the following: <https://www.icann.org/en/system/files/correspondence/namazi-to-wei-et-al-15feb19-en.pdf> which states:

*The CJK GPs correctly note that visual similarity of strings causing confusability should be resolved by String Similarity Panels during TLD string evaluation, String Confusion Objection and String Contention Mechanisms. While recognizing this, the Procedure additionally states that “the LGR process is designed to clear the table of all the straightforward, non-subjective cases, mainly by returning a ‘blocked’ disposition," further explaining that “Even for variants based on visual similarity, there exists a subset of evaluation rules that could be applied in an automated manner, obviating the need for further case-by case or even contextual review.”*

*Because the Integration Panel (IP) is governed by the Procedure, it has asked every GP to provide a list of visually identical code points or sequences, limited to the cases which are “straightforward and non-subjective”, to be treated as blocked variants. All GPs up to date have undertaken the analysis and provided a list of the candidates determined. This is an important part of the Procedure and helps ensure that the RZ-LGR satisfies the “Contextual Safety Principle” which asks to minimize the risk of malicious use of IDN labels5 . As well noted in your communication, the visually similar strings, which are not deemed identical, will continue to be evaluated by the String Similarity Review processes. Due to the independent nature of the working of the IP, as stipulated by the Procedure, ICANN org suggests that the CJK GPs continue discussions with the IP in order to find a mutually agreeable solution based on the Procedure. ICANN org will organize a face to face meeting of the GPs and IP at the ICANN64 meeting in Kobe, Japan to continue to support this dialogue.*

# References

[Procedure] Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels: <https://www.icann.org/en/system/files/files/draft-lgr-procedure-20mar13-en.pdf>

[OutOfRepertoireVariants] Out of Repertoire Variants in Root-Zone LGR and Proposals: <https://www.icann.org/en/system/files/files/root-zone-lgr-repertoire-variants-25sep17-en.pdf>

1. This pair is not (yet) in the Unicode confusable list, but is part of dotAsia. [↑](#footnote-ref-0)
2. This pair is not (yet) in the Unicode confusable list, but is similar to other pairs in that list (it is already a CLGR13 variant set and would show as a grey highlighted pair above) [↑](#footnote-ref-1)