Comments on N2225 (CJK Supplementary Components for IDS Use), plus Suggestions for Additional List of Components and New Additional Components

Author: Leroy Vargas
Type: Individual Contribution to IRG #49
Date: 2017/08/27
References: IRG N1183; IRG N2117; IRG N2204; IRG N2225 Feedback (Additional Components from Lee Collins; SAT Comments); IRG N2242; IRG N2243; Unicode Standard Annex #45

Comments on SAT's individual issues:

On 2017/07/24, SAT raised the issue of components that can be IDS-decomposed into other sub-components (general principle 1.c in their comments), of which some or all are already encoded as CJK ideographs. This may be true, but only in the context of deep, maximal IDS decomposition, which results in a sometimes very lengthy IDS sequence that may be easy for machine checking, but hard for eye-ball checking or for checking whether two similar sequences are cognate/non-cognate and/or share/differ in phonetic value.

Principle 2.1 of IRG N1183 (Minimal division) recommends not to "divide too much", stressing that further division can be easily achieved through machine recursion. The principal idea behind N2204 by Henry CHAN and N2225 by Chan and Prof. Qin LU is to improve readability and quality of IDS sequences – which may imply minimizing the lengths of IDS sequences (by employing as few IDCs as possible). Also, further decomposing a decomposable component may do away with the overall component's phonetic value –and thus, the intended meaning–, thus making it harder to check whether two similar components with slightly-different subcomponents but potentially equal phonetic value might share the same meaning and thus be unified or not. So, while it is technically desirable to single out a subcomponent of a complex component by IDS decomposition so that it can be reused by other components, in practice it is also desirable to keep the overall component with a fixed phonetic and/or meaning so as to optimize human-readable IDS generation, especially when two or more components sharing the same subcomponent are semantically related to each other in the context of variance or make up a traditional/simplified pair (or, in some instances, a traditional/J-simplified/C-simplified trio, or a traditional/C-simplified/V-simplified trio, etc.).

Hereby, the author of this feedback document suggests instead splitting N2225's list of unencoded supplementary components (currently section *"2. List of Components"*, aliased §2 from here on) into two sections: one listing non-IDS-decomposable components (called atomic components) and another one listing IDS-decomposable components (called non-atomic components) that may or may not contain any of the atomic subcomponents from the first section. §2 can be renamed to *"2. List of Atomic Components"* while current section *"3. Additional Components for Reference"* (§3o) can be renumbered to *"4. Additional Components for Reference"* (§4); the new list of decomposable components can become new section *"3. List of Non-Atomic Components"* (§3n to distinguish it from §3o which is itself renamed to §4), as can be previewed at the near end of this document.

As for the individual issues: consider SAT's first one concerning components H9-01, H14-01, & S5-01 shown below. SAT argued that rather than proposing these three separate components, it was enough to merely single out the right-hand subcomponent of both H9-01 & S5-01, which is also the top-right subcomponent of H14-01, as a proposed component (to be given IDS code P3-04 by this author) that could be reused by other components via IDS composition. However, following the same decomposition logic suggested by SAT, it appears that the same subcomponent P3-04, circled in red below, can be itself decomposed into 日乍一 (or its PRC/Vietnamese variant 日乍丶, as used in S5-01). (Its traditional form found in H14-01 and circled green below, which is given here IDS code P8-03, also contains P3-04 and decomposes into 日乍一罒.)

IDS Code	Glyph	Description
H9-01	臦	CJK COMPONENT SIMPLIFIED OVERSIGHT Example: top of 監 / 覧 / 竪 = CDP-8CC9
H14-01	臨	CJK COMPONENT OVERSIGHT Example: top of 鑒 /覽 = CDP-8D50
S5-01	川大	CJK COMPONENT C-SIMPLIFIED OVERSIGHT Example: top of 監 / 览 = CDPO-81F0 / UTC-02965

However, further decomposing P3-04 would hide the fact that it can assume at least two variant shapes, each differing by the type of stroke at the bottom ('一' versus '丶', both which are non-cognates and thus cannot be individually unified), so it is best to keep P3-04 as an "atomic component" in spite of its obvious decomposition(s). Also, it is in the best interest to keep P8-03 as an atomic component in spite of the presence of P3-04 in it, because both P8-03 & P3-04 appear to form a traditional/simplified pair, just like 釒 & 钅; they can both be allocated into §2 like this:

IDS Code	Glyph	Description
P3-04	乍 乍	CJK COMPONENT SIMPLIFIED PART OVERSIGHT Example: top right corner of 監 / 覧 Component proposed by SAT Variant form is for PRC, also requested for Vietnamese by Lee COLLINS
P8-03	盇	CJK COMPONENT PART OVERSIGHT Example: top right of 鑒 /覽 New component proposed by this author

Extending the same reasoning to the original components H14-01, H9-01, & S5-01, which all share a common meaning ("oversight") and a respective traditional/J-simplified/C-simplified form relationship, the author of this feedback believes that rather than just deleting these three components for the mere reason that each contains P3-04, they can simply be relocated to §3n, with their respective minimal IDS decompositions ⿰臣&P8-03;, ⿰臣&P3-04;, & ⿰&S2-01;&P3-04; recorded alongside for machine-recursive decomposition purposes.

In another issue, SAT also pointed out the decomposability of S4-01, which happens to make use of S2-01, already present in §2 and also used by S5-01 above. Following the same logic laid out above, this author also suggests moving S4-01 to §3n (with decomposition ⿰&S2-01;又).

IDS Code	Glyph	Description
S4-01	刂又	CJK COMPONENT C-SIMPLIFIED LEAD Example: top of 坚 / 竖 / 紧 / 肾 / 贤 = CDPO-81EF / UTC-02964

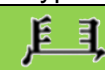SAT also pointed out the decomposability of P5-01, singling out its left-half subcomponent as a candidate proposal (this author suggests IDS code P3-05, originally treated by N1183's appendix text file IRGN1183A_ids_irg.txt as a variant of U+2008E 刂), but did not provide evidence of other ideographs using P3-05 to support such singling-out. Nevertheless, this author concurs with SAT in adding P3-05 to §2 and relocating P5-01 to §3n (with decomposition ⿰&P3-05;刀 for recursion purposes).

IDS Code	Glyph	Description
P5-01	卬刀	CJK COMPONENT LIU Example: top of 留 / top left of 劉 = CDP-8C69

IDS Code	Glyph	Description
P3-05	卬	CJK COMPONENT PART LIU Example: top left of 留 Component proposed by SAT

Concerning P15-01, SAT suggested "making a component without 同 in the middle" (to be given here IDS code P9-01), but again did not provide evidence of ideographs using P9-01 but not 同. So, although it may be a good idea to propose P9-01 at §2 (in the same spirit as the similarly-looking P10-01), this author still suggests moving P15-01 to §3n (with decomposition ⿲&P9-01;同) rather than dismissing it.

IDS Code	Glyph	Description
P15-01	𦥑	CJK COMPONENT COMPLEX POURING VESSEL Example: top of 釁 / 爨

IDS Code	Glyph	Description
P9-01	𦥑	CJK COMPONENT PART POURING VESSEL Example: top enclosure of 釁 / 爨 New component proposed by SAT

Comments on SAT's proposal for addition:

The two components proposed for addition by SAT, using the same logic used by SAT to question many of the decomposable components like those mentioned above, can also be decomposed into already-encoded subcomponents, but since they do not exist as standalone encoded ideographs and they seem to have an implied semantic and/or phonetic value, they can too be listed in §3n with IDS codes, decompositions, and suggested provisional names given below:

3.1.   FS = 1 (FS Code = H)

IDS code	Glyph	Decomposition	Description
H4-02	士	⿱十冖	CJK COMPONENT COMET TOP Example: top of 索, 孛, 寁 etc. = CDP-8D52

3.4.   FS = 4 (FS Code = D)

IDS code	Glyph	Decomposition	Description
D4-03	立	⿱亠丷	CJK COMPONENT ANCIENT FAULT TOP Example: top of 辛, 帝, 商, 啇, 产 etc. = CDP-8BAE

Comments on additional components proposed by Lee Collins:

On 2017/07/21, Lee COLLINS suggested to add 17 new components to the list originally formulated in N2225 §2. Following the same logic used by SAT above, 11+ of these components can be decomposed into smaller subcomponents (of which one is already proposed as standalone ideograph in IRG Working Set 2015 and another one is proposed new by this author [although it might be unifiable with a similar, already-encoded ideograph]); also, one component is already proposed standalone in WS 2015, and another one is a variant of an already-encoded ideograph:

2.1.   FS = 1 (FS Code = H)

IDS code	Glyph	Description + comments, suggestions, & additions by Vargas
H9-02	聂	CJK COMPONENT 78CA VARIANT ONE WS2015 sequence number 02833 (UTC-01039), can be decomposed into ⿱石双. Including this as a component into N2225 list(s) might seem questionable or at best debatable (Does this belong to §3n or §4?)
H9-03	㪠	CJK COMPONENT 8944 VARIANT ONE Example: bottom right of 爆 Can be decomposed into ⿱共仦, suggest moving to §3n and renaming it CJK COMPONENT HELP VARIANT ONE.
H11-01	莫	CJK COMPONENT 6F22 RIGHT HALF Example: right of 漢, unifiable variant of U+26C29 𦰩 Suggest moving this to §4 as an "additional component for reference".
H14-02	㘆	CJK COMPONENT 21570 VARIANT ONE Example: right of 钁 Can be decomposed into ⿱⿲罒凵米八夂, suggest moving to §3n and renaming it CJK COMPONENT ZONG VARIANT ONE.

2.2. FS = 2 (FS Code = S)

IDS code	Glyph	Description + comments, suggestions, & additions by Vargas
S4-03	듁	CJK COMPONENT 2A7FD VARIANT Example: right of 旳 Can be decomposed into 日卜丂, suggest moving to §3n.
S8-02	罪	CJK COMPONENT 7F85 VARIANT ONE Example: left of 罜, right of 緋 Can be decomposed into 日罒井, suggest moving to §3n and renaming it CJK COMPONENT GAUZE VARIANT ONE.
S8-03	罗	CJK COMPONENT 7F85 VARIANT TWO Example: right of 鋰, bottom of 蕁 Can be decomposed into 日罒丱, suggest moving to §3n and renaming it CJK COMPONENT GAUZE VARIANT TWO.
S8-04	烍	CJK COMPONENT RAISE THREE Example: right of 噗 Enclosure of S11-02; similar to S7-03; maybe unifiable w/U+2D048 㵈?
S11-02	犇	CJK COMPONENT 8209 VARIANT Example: right of 嚹 Can be decomposed into two unencoded components: ☐⑧③, where ⑧ is the component proposed above as S8-04 (unifiable with U+2D048 㵈?) and ③ is WS2015 sequence num. 00002 (UTC-01671), which is similar to U+20006 千; suggest moving this component to §3n and renaming it to CJK COMPONENT RAISE VARIANT.

2.3. FS = 3 (FS Code = P)

IDS code	Glyph	Description + comments, suggestions, & additions by Vargas
P4-06	�	CJK COMPONENT 98ED RIGHT HALF Example: right of 飭 Can be decomposed into 日𠂉力, suggest moving to §3n and renaming it CJK COMPONENT ORDER RIGHT HALF.
P8-02	舀	CJK COMPONENT 8200 VARIANT Example: right of 陷 Can be decomposed into 日爫旧, suggest moving to §3n and renaming it CJK COMPONENT LADLE VARIANT.

2.4. FS = 4 (FS Code = D)

IDS code	Glyph	Description + comments, suggestions, & additions by Vargas
D8-01	姜	CJK COMPONENT 7FA9 VARIANT ONE E.g: right of 僟, 犠, 議 Can be decomposed into 曰丷又, suggest moving to §3n and renaming it CJK COMPONENT RIGHTEOUS VARIANT ONE.
D9-01	其	CJK COMPONENT TRIGRAM FIVE VARIANT ONE E.g: left of 辮 Can be decomposed into 曰丷共, suggest moving to §3n.
D13-01	興	CJK COMPONENT 8208 VARIANT ONE Example: left of 觪, bottom of 轟 Can be decomposed into 曰回&D4-01;同&H3-04;, where D4-01 is being repurposed as an enclosing component (see author's comment on such component two pages below); suggest moving to §3n & renaming it to CJK COMPONENT PROSPER VARIANT ONE.

Also, this author identified three non-decomposable components by Collins whose provisional hex-numeric names could be renamed to more meaningful ones:

2.3. FS = 3 (FS Code = P)

IDS Code	Glyph	Description + comments, suggestions, & additions by Vargas
P3-03	个	CJK COMPONENT 4E2A VARIANT ONE Example: right of 斸 Derived from 个, not 4E87 Suggest renaming it CJK COMPONENT SINGLE VARIANT ONE.

2.4. FS = 4 (FS Code = D)

IDS Code	Glyph	Description + comments, suggestions, & additions by Vargas
D4-02	宠	CJK COMPONENT 5198 VARIANT ONE E.g: bottom right of 㧣 Suggest renaming it CJK COMPONENT MOVE ON VARIANT ONE.
D5-03	宠	CJK COMPONENT 5198 VARIANT TWO E.g: right of 挽 Suggest renaming it CJK COMPONENT MOVE ON VARIANT TWO.
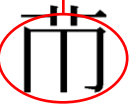
Comments on some original components + author's own additions:

The author of this feedback document is also adding his own comments and suggestions to some decomposable components that may have been overlooked by SAT and other involved parties, along with proposing some new additions of his own after analyzing some of the decomposables in below and comparing them to other already-encoded ideographs that contained the same component(s):

2.1. FS = 1 (FS Code = H)

IDS code	Glyph	Description + comments, suggestions, & additions by Vargas
H2-01	与	CJK COMPONENT SIMPLIFIED AND Example: enclosure of 与, middle enclosure of 鳥 Similar in shape to U+20001 ㄅ, but not cognate; enclosure of H6-01.
H3-04	六	CJK COMPONENT RAISE ONE Example: bottom of 具 / 巺 / 共 Can be decomposed into 日一八, suggest moving to §3n.
H6-01	烏	CJK COMPONENT BEAST BODY TWO Example: bottom of 鷹 / 焉 / 鳥 = CDP-8DBF Can be decomposed into □②灬, where ② is a new component to be proposed as H2-01; suggest creating new component H2-01 above at §2 and moving this overall component to §3n.
H6-02	厽	CJK COMPONENT LIGATURE DOUBLE ARRIVE Example: top of 晉 = CDP-8DAB Can be decomposed into 日一厶一, suggest moving to §3n.

2.2. FS = 2 (FS Code = S)

IDS Code	Glyph	Description + comments, suggestions, & additions by Vargas
S3-01	冂冂	CJK COMPONENT ANCIENT PAIR Example: enclosure of 网, bottom enclosure of 繭 Bottom part of S7-01.
S7-01	苗	CJK COMPONENT COCOON Example: outer enclosure of 繭 / 蠒 Can be decomposed into either 日卄③ or 日丗③, where ③ is a new component to be proposed as S3-01; suggest creating new component S3-01 above at §2 and moving this overall component to §3n.
S11-01	茜 茜 茜	CJK COMPONENT DREAMING PERSON Example: top of 夢 / 薨 / 瞢 / 薂 / 薨 / 蔓 / 薲 = CDP-8D60 Can be decomposed into either 日卄罒宀 or 日丗罒宀 (depending on desired variant), suggest moving to §3n.

2.3.   FS = 3 (FS Code = P)

IDS Code	Glyph	Description + comments, suggestions, & additions by Vargas
P2-02	𡿨	CJK COMPONENT OMISSION RIGHT HALF Example: top right of 楽 / 興(D13-01) Right-half subcomponent of D4-01, especially when used as enclosure.
P6-01	甸	CJK COMPONENT BEAST BODY ONE Example: bottom of 舄 = CDP-8C40 Can be decomposed into 囗勹灬, suggest moving to §3n.
P6-03	㣺	CJK COMPONENT COMPLEX PERSON CAVE Example: top of 奂 / 夐 = CDP-8CAF Basic form can be decomposed into 囗勹冈; variant forms must check whether bottom-most component shows same variation in isolation. Anyway, suggest moving to §3n.

2.4.   FS = 4 (FS Code = D)

IDS Code	Glyph	Description + comments, suggestions, & additions by Vargas
D4-01	⽶	CJK COMPONENT OMISSION Example: bottom of 裹 / bottom right 涘 / enclosure of 楽 / 興 / 屄(尾) = CDP-89AE Can be decomposed into either 囗丶② (where ② is a new component to be proposed as P2-02) or 囗二二 (TW & HK variant form); suggest creating P2-02 above at §2 and moving this overall component to §3n. This author also suggests using this component as an enclosure component; please see comment on Collins' D13-01 above.
D5-02	𭕄	CJK COMPONENT C-SIMPLIFIED LEARN Example: top of 学 / 喾 / 栄 / 鸴 / 鲎 / 覚 / 蛍 etc. = CDPO-81DD / UTC-02966 Can be decomposed into 囗丷冖, suggest moving to §3n.
D7-01	亯	CJK COMPONENT ARCHITECTURAL STRUCTURE Example: top of 亮 / 亭 = CDP-8C4D Can be decomposed into 囗亠口冖, suggest moving to §3n.
D7-02	宷	CJK COMPONENT GUEST Example: top of 賓 / 賓, top left of 叡 = CDP-8D74 Decomposable into 囗宀一少 (囗宀一少 is a variant), suggest moving to §3n. The bottom subcomponent can be also spun off below as a new, decomposable component H4-01 that could also be added to §3n.

3.1.   FS = 1 (FS Code = H)

IDS code	Glyph	Decomposition	Description
H4-01	丕	囗一少 囗一少	CJK COMPONENT PART GUEST Example: middle of 賓 / 賓, right half of 桻, bottom of �White, bottom left of 頻, middle of 鬪 New component proposed by this author

This author also suggests synthesizing a new enclosing component, P4-07, by combining together P2-01 and Z2-01 which are often (but not exclusively) used in pairs. P4-07 should be put into §3n.
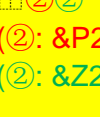
2.3. FS = 3 (FS Code = P)

IDS Code	Glyph	Description + additions by Vargas
P2-01		CJK COMPONENT LEFT ENCLOSURE Example: left of 卯, top left of 兜

2.5. FS = 5 (FS Code = Z)

IDS Code	Glyph	Description
Z2-01		CJK COMPONENT RIGHT ENCLOSURE Example: right of 兜 / top right of 叚

3.3. FS = 3 (FS Code = P)

IDS code	Glyph	Decomposition	Description
P4-07		⿻②② (②: &P2-01;) (②: &Z2-01;)	CJK COMPONENT ENCLOSURE Example: top enclosure of 兜 / 兜 New component proposed by this author

Additional new components proposed by this author:

In addition to the new components proposed above (as a response to comments by SAT, proposals by Collins, and as the result of further analysis of some of the original components proposed by Chan & Lu), this author adds below a few more new components taken from other encoded ideographs, including two decomposable components:

2.2. FS = 2 (FS Code = S)

IDS Code	Glyph	Description
S7-04	里	CJK COMPONENT YELLOW LOAM Example: bottom of 菫 / 堇

2.3. FS = 3 (FS Code = P)

IDS Code	Glyph	Description
P7-04	◇◇◇	CJK COMPONENT 20137 TOP Example: top of 胥

2.5. FS = 5 (FS Code = Z)

IDS Code	Glyph	Description
Z3-04	㐅	CJK COMPONENT DREAD Example: bottom of 喪 / 畏 Note: similar to 仪

3.1. FS = 1 (FS Code = H)

IDS code	Glyph	Decomposition	Description
H5-02	壳	⿱士冖	CJK COMPONENT SHELL ONE Example: top of 壳, top left of 殼 / 毀 = NUCV-264a form 1
H6-04	壴	⿱士冖一	CJK COMPONENT SHELL TWO Example: top of 壳, top left of 殼 / 毀 = NUCV-264a form 2

This author also adds to §4 (formerly §3o) an additional component to be provisionally used "for reference", until UTC-00984 (sequence number 00470 in IRG Working Set 2015) gets encoded into the UCS:

Base Character	Variant Glyph	Descriptions
U+9FB4 ⺆	⺆	Example: bottom of 韦 / 令 (J-source) = UTC-00984

Consolidation of all comments & ==additions== above:

This author, as hinted at the beginning of this feedback document, suggests creating a new list of unencoded decomposable components for IDS use, in order to accommodate many of the components questioned by SAT and most of the new additions suggested by Collins, along with relocated original components that may fit into this category (based on their decomposition) and ==most of this author's own proposed additions==.

As stated in the first page of this document, IRG N2225's current section *"2. List of Components"* is suggested to be renamed to section *"2. List of ==Atomic== Components"* so that it only lists unencoded components that cannot be resolved into IDS sequences of simpler subcomponents; some of SAT's own atomic (and quasi-atomic) component suggestions may also be added to this renamed list, along with ==some additions suggested by this author==. Old section *"3. Additional Components for Reference"* is to be pushed down as renumbered section *"==4==. Additional Components for Reference"* in order to accommodate the new proposed list; in addition to its currently-contained variants, one of the Collins suggestions questioned above fits into this field.

This author's new list becomes new section *"==3. List of Non-Atomic Components=="*, and it accommodates unencoded components that can be decomposed into other subcomponents. §3n is to follow the same layout as §2: the list is sub-partitioned according to first stroke. In order to avoid changing components' IDS codes, components from §2 that may be moved into this new section will keep the same IDS code, analogous to Unifiable-Component-Variant (UCV) & Non-Unifiable-Component-Variant (NUCV) rules in IRG Working Document Series (IWDS) documents N2242 & N2243 that maintain their rule numbers even if they switch from one list to another. Decompositions are shown in a ==new column== inserted between the glyph and description columns.

In order to better visualize the new section, the author has decided to consolidate N2225's original tables with the suggested additions (==Collins==, ==SAT==, ==Vargas==) and relocations in the following pages:

1. Introduction

FS	FS Code	Name of Stroke	Rep Glyph	Character Examples
1	**H**	Horizontal stroke ("橫", *Héng*)	一	一 (one), 三 (three)
2	**S**	Vertical stroke ("豎", *Shù*)	丨	卜 (divination), 巾 (thirty)
3	**P**	Left slant stroke ("撇", *Piě*)	丿	川 (river), 月 (moon)
4	**D**	Dot stroke ("點", *Diǎn*)	、	流 (current), 宝 (precious)
5	**Z**	Turned stroke ("折", *Zhé*)[1].	乙 ∟	㣇, 予, 細

---

[1] There are many different forms of the turned strokes and their starting stroke can be any of the above single strokes. No distinction is made in this document (IRG N2225).

2. List of Atomic Components
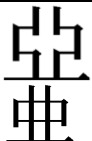
2.1. FS = 1 (FS Code = H)

IDS code	Glyph	Description
H2-01		CJK COMPONENT SIMPLIFIED AND Example: enclosure of 与, middle enclosure of 舄
H3-01		CJK RADICAL PAW FOUR[2] Example: bottom of 虐 / 疟, bottom of 虐 / 仓 = CDP-8DC1 / UTC-00969 (variant glyph = USAT-05803; may be confusable with P3-01 below)
H3-02		CJK COMPONENT C-SIMPLIFIED TRIPLE EARTH Example: 尧 = CDP-89D9
H3-03		CJK COMPONENT CAI Example: outside of 在 / 存 = CDP-88F1
H5-01		CJK COMPONENT SACK DRAWSTRING WITH STOPPER Example: top of 棗 and 囊 = CDP-8DDD
H6-03		CJK COMPONENT PART HORSE Example: outside of 馬 = CDP-896A
H7-01		CJK COMPONENT LIGATURE DOUBLE EAST Example: top of 曹 / 薔 = CDP-8D61
H8-01		CJK COMPONENT LEAF Example: top of 募 / 憂 / 戛 = CDP-8CD4
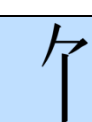
2.2. FS = 2 (FS Code = S)

IDS Code	Glyph	Description
S2-01		CJK COMPONENT C-SIMPLIFIED MOUND Example: left of 帅 / 归 = CDP-8968
S3-01		CJK COMPONENT ANCIENT PAIR Example: enclosure of 网, bottom enclosure of 爾
S4-02		CJK COMPONENT BOTTOMLESS HOLLOW Example: bottom of 隽 = CDP-8DDF
S5-02		CJK COMPONENT REPTILE BODY Example: body of 竜, 龟 = CDP-8BE8

---

[2] The name of the components used in this document (IRG N2225) is only meant as a reference. It has no significance in the IDS sequence. So, its appropriateness will not be discussed for WS 2017. However, if the naming of these components is changed as a result of IRG/WG discussion, update to this document (IRG N2225) will be done accordingly.

IDS Code	Glyph	Description
S7-02		CJK COMPONENT VASE BOTTOM Example: bottom of 壷 / 壺 / 壹 / 歡
S7-03		CJK COMPONENT 2E5D3 TOP Example: top of 覺
S7-04		CJK COMPONENT YELLOW LOAM Example: bottom of 菫 / 菫
S8-01		CJK COMPONENT DRAGON BODY Example: right of 龍 = CDP-89B0
S8-04		CJK COMPONENT RAISE THREE Example: right of 㷠; might be unifiable with U+2D048 㹈?
S9-01		CJK RADICAL J-SIMPLIFIED FROG Example: right of 縄 /蝇

2.3. FS = 3 (FS Code = P)

IDS Code	Glyph	Description
P2-01		CJK COMPONENT LEFT ENCLOSURE Example: left of 卬, top left of 兜
P2-02		CJK COMPONENT OMISSION RIGHT HALF Example: top right of 樂 /猴
P3-01		CJK RADICAL PAW THREE Example: left of 印 / 袠 / 裒 / 褭 / 褺 / 虜
P3-03		CJK COMPONENT SINGLE VARIANT ONE Example: right of 卧 Derived from 个, not 4E87
P3-04		CJK COMPONENT SIMPLIFIED PART OVERSIGHT Example: top right corner of 監 / 覽
P3-05		CJK COMPONENT PART LIU Example: top left of 留
P4-01		CJK RADICAL V-SIMPLIFIED WIND Example: right of 颵
P4-02		CJK RADICAL MEAT TWO Example: top of 炙 / 磊 / 然 /將
P4-03		CJK RADICAL MOON TWO Example: top right of 望 = CDP-8B5E
P4-04		CJK COMPONENT LIGATURE DOUBLE HUMAN Example: bottom right of 旅 (J-source) / right of 尓

P4-05	鸟	CJK COMPONENT SIMPLIFIED PART BIRD Example: outside of 鸟 / 岛
P5-02	𰀲	CJK COMPONENT LOOSE ROCK ONE Example: left of 段 = CDP-8CE4
P6-02	龟	CJK COMPONENT BEAST HEAD ONE Example: top of 象 / 烏
P7-01	𦥯	CJK COMPONENT FOOD VESSEL TWO Example: left of 卽(+VS19: 卽) / 既(+VS20: 既)
P7-02	𡊬 肎 用	CJK COMPONENT BEAST BODY THREE Example: bottom of 劚 / 壏 / 簋 / 鎘 / 甪 / 鼠 / 攛
P7-03	鳥	CJK COMPONENT PART BIRD Example: outside of 島 / 鳥 = CDP-8CBB
P7-04	◇◇◇	CJK COMPONENT 20137 TOP Example: top of 𥄂
P8-01	無	CJK COMPONENT DANCE Example: top of 無 / 舞 = CDP-8BBE
P8-03	𥄂	CJK COMPONENT PART OVERSIGHT Example: top right of 鑒 /覽
P9-01	𦥑	CJK COMPONENT PART POURING VESSEL Example: top enclosure part of 爨 / 爨
P10-01	𦥯	CJK COMPONENT RAISE TWO Example: enclosure of 與 / 与 / 興 / 輿 = CDP-8BA8

2.4. FS = 4 (FS Code = D)

IDS Code	Glyph	Description
D4-02	宠	CJK COMPONENT MOVE ON VARIANT ONE Example: bottom right of 𢪈
D5-03	宠	CJK COMPONENT MOVE ON VARIANT TWO Example: right of 抚
D6-01	亦	CJK COMPONENT C-SIMPLIFIED CHAOS Example: top of 恋 / 娈 / 峦. Note: use 4EA6 亦 instead where appropriate, such as 5955 奕 = UTC-02967

D6-02	�	CJK COMPONENT VIRTUOUS Example: left of 刖 / 朗, left hand side form of 良 = CDP-8C42
D6-03	声	CJK COMPONENT BEAST HEAD TWO Example: outside of 鹿 / 鷹 = CDP-8D56
D10-01	㳄	CJK COMPONENT EVEN Example: top of 𪎭 / 齊 = CDP-8DEB

2.5. FS = 5 (FS Code = Z)

IDS Code	Glyph	Description
Z2-01	⌐	CJK COMPONENT RIGHT ENCLOSURE Example: right of 兜 / top right of 叚
Z2-02	丩	CJK COMPONENT LEFT HAND ONE Example: left of 収
Z2-03	𠃊	CJK COMPONENT LEFT HAND TWO Example: right of 虺 = CDP-88B5
Z2-04	马	CJK COMPONENT SIMPLIFIED PART HORSE Example: outside of 马
Z2-05	又	CJK COMPONENT RIGHT HAND TWO Example: right top of 祭 / 登 = CDP-8C4E
Z3-01	彐 彐	CJK COMPONENT RIGHT HAND ONE Example: top of 尋 / top right of 侵 / bottom of 彗 = CDP-8BAB
Z3-02	丮	CJK COMPONENT HOLD ONE Example: right top of 恐 / 鞏 etc.
Z3-03[3]	丸	CJK COMPONENT HOLD TWO (variant of 丮) Example: right of 執 / 埶 / 㹙 / 秇 Note: use 4E38 丸 (wan2) where appropriate, e.g. 7D08 紈.
Z3-04	𢆶	CJK COMPONENT DREAD Example: bottom of 喪 / 畏 Note: similar to 仸
Z4-01	𠂢	CJK COMPONENT TILTED HEAD Example: bottom of 吳 = CDP-8CD5
Z4-02	丮	CJK COMPONENT RAN Example: left of 那 = CDP-89FA

---

[3] Refer to IRG N2225 Appendix A: Rationale for Inclusion of Z3-03 (reprinted in this feedback just in case).

Z5-01	�==(glyph)	CJK COMPONENT FOOD VESSEL ONE Example: left of 即 / 既 = CDP-8B7C
Z5-02	(glyph)	CJK COMPONENT LOOSE ROCK TWO Example: left of 叚 = CDP-8C7A
Z7-01	(glyph)	CJK COMPONENT BEAST BODY THREE Example: bottom of 鼠 / 鼵(TW/HK variant glyph) = CDP-8D46

3. List of Non-Atomic Components

3.1.   FS = 1 (FS Code = H)

IDS code	Glyph	Decomposition	Description
H3-04	(glyph)	⿱一八	CJK COMPONENT RAISE ONE Example: bottom of 具 / 巺 / 共
H4-01	(glyph)	⿱一少 ⿱一少	CJK COMPONENT PART GUEST Example: middle of 賓 / 賓, right half of 梥, bottom of 岑, bottom left of 頻, middle of 闚
H4-02	(glyph)	⿱十宀	CJK COMPONENT COMET TOP Example: top of 索, 字, 寁 etc. = CDP-8D52
H5-02	(glyph)	⿱士宀	CJK COMPONENT SHELL ONE Example: top of 壳, top left of 殼 / 殻 = NUCV-264a form 1
H6-01	(glyph)	⿱&H2-01;灬	CJK COMPONENT BEAST BODY TWO Example: bottom of 廌 / 焉 / 鳥 = CDP-8DBF
H6-02	(glyph)	⿱一厸一	CJK COMPONENT LIGATURE DOUBLE ARRIVE Example: top of 晉 = CDP-8DAB
H6-04	(glyph)	⿱士宀一	CJK COMPONENT SHELL TWO Example: top of 壳, top left of 殼 / 殻 = NUCV-264a form 2
H9-01	(glyph)	⿰臣&P3-04;	CJK COMPONENT SIMPLIFIED OVERSIGHT Example: top of 監 / 覽 / 竪 = CDP-8CC9
H9-03	(glyph)	⿱圭衣	CJK COMPONENT HELP VARIANT ONE Example: bottom right of 爆
H14-01	(glyph)	⿰臣&P8-03;	CJK COMPONENT OVERSIGHT Example: top of 鑒 /覽 = CDP-8D50
H14-02	(glyph)	⿱⿰凵米八夊	CJK COMPONENT ZONG VARIANT ONE Example: right of 钀

3.2. FS = 2 (FS Code = S)

IDS code	Glyph	Decomposition	Description
S4-01	收	⿰&S2-01;又	CJK COMPONENT C-SIMPLIFIED LEAD Example: top of 坚 / 竖 / 紧 / 肾 / 贤 = CDPO-81EF / UTC-02964
S4-03	丂	⿱卜丂	CJK COMPONENT 2A7FD VARIANT Example: right of �
S5-01	际	⿰&S2-01;&P3-04;	CJK COMPONENT C-SIMPLIFIED OVERSIGHT Example: top of 监 / 览 = CDPO-81F0 / UTC-02965
S7-01	帯	⿱卝&S3-01; ⿱卝&S3-01;	CJK COMPONENT COCOON Example: outer enclosure of 繭 / 蠒
S8-02	罘	⿱罒卝	CJK COMPONENT GAUZE VARIANT ONE Example: left of 脞, right of 絣
S8-03	罘	⿱罒屮	CJK COMPONENT GAUZE VARIANT TWO Example: right of 鑼, bottom of 蕐
S11-01	茻	⿱卝罒冖 ⿱卝罒冖	CJK COMPONENT DREAMING PERSON Example: top of 夢 / 薨 / 瞢 / 蔑 / 薨 / 覆 / 薨 = CDP-8D60
S11-02	举	⿱⑧③ (⑧: &S8-04; 兴) (③: UTC-01671)	CJK COMPONENT RAISE VARIANT Example: right of 嘩

3.3. FS = 3 (FS Code = P)

IDS code	Glyph	Decomposition	Description
P4-06	夯	⿱𠂉力	CJK COMPONENT ORDER RIGHT HALF Example: right of 飭
P4-07	𫂁	⿰②② (②: &P2-01;) (②: &Z2-01;)	CJK COMPONENT ENCLOSURE Example: top enclosure of 兜 / 兜
P5-01	夗	⿰&P3-05;刀	CJK COMPONENT LIU Example: top of 留 / top left of 劉 = CDP-8C69
P6-01	匒	⿰勹灬	CJK COMPONENT BEAST BODY ONE Example: bottom of 舃 = CDP-8C40

17

IDS code	Glyph	Decomposition	Description
P6-03		⿱勹內	CJK COMPONENT COMPLEX PERSON CAVE Example: top of 奐 / 奐 / 夐 = CDP-8CAF
P8-02		⿱⺌臼	CJK COMPONENT LADLE VARIANT Example: right of 陷
P15-01		⿱&P9-01;同	CJK COMPONENT COMPLEX POURING VESSEL Example: top of 爨 / 爨

3.4.   FS = 4 (FS Code = D)

IDS code	Glyph	Decomposition	Description
D4-01		⿰冫&P2-02; ⿱二二 (TW/HK)	CJK COMPONENT OMISSION Example: bottom of 褭 / 澁, enclosure of 楽 / 燊 / 屄(尾) = CDP-89AE
D4-03		⿱亠丷	CJK COMPONENT ANCIENT FAULT TOP Example: top of 辛, 帝, 商, 啇, 产 etc. = CDP-8BAE
D5-02		⿱⺍冖	CJK COMPONENT C-SIMPLIFIED LEARN Example: top of 学 / 嚳 / 燅 / 鸴 /鱟 / 觉 / 蝿 etc. = CDPO-81DD / UTC-02966
D7-01		⿱亠口冖	CJK COMPONENT ARCHITECTURAL STRUCTURE Example: top of 亮 / 亭 = CDP-8C4D
D7-02		⿱宀&H4-01;	CJK COMPONENT GUEST Example: top of 賓 / 賔, top left of 叡 = CDP-8D74
D8-01		⿱羊又	CJK COMPONENT RIGHTEOUS VARIANT ONE Example: right of 僕, 㩦, 謢
D9-01		⿱丷共	CJK COMPONENT TRIGRAM FIVE VARIANT ONE Example: left of 瀷
D13-01		⿱⿲④同③ (④: &D4-01;) (③: &H3-04;)	CJK COMPONENT PROSPER VARIANT ONE Example: left of 鱦, bottom of 蠱

3.5.   FS = 5 (FS Code = Z) (currently none allocated here)

18

4. Additional Components for Reference

Base Character	Variant Glyph	Descriptions
U+9FB4 マ	卩	Example: bottom of 靑 / 令 (J-source) = UTC-00984
U+2B94E 兔	冬	Example: right of 終 / 鮗
U+53C0 専	叀	Example: top of 專 / 惠 = CDP-8BD0
U+26C29 莫	莫	Example: right of 漢
U+5C6F 屯	七	Example: right of 扗 / 酕 / 鈍 = CDP-8AF0
U+6B62 止	㐫	Example: bottom of 歴 / 歷 = CDP-8962
U+9FB0 ㄓ	之	Example: bottom of 走 / 夋 / 縱 = CDP-8961
U+3B30 叓	叓	Example: right of 悙 / 㻕 / 諶 = CDP-89E7
U+4E23 亝	亝 亝	Example: top right of 猶, right of 聇 / 颵 / 騧 = CDPO-85FD

5. Questionable Components for Further Discussion

IDS code	Glyph	Description
H9-02	叜	CJK COMPONENT 78CA VARIANT ONE The author of this feedback cannot see how to fit this Collins-proposed component (which was already submitted to IRG Working Set 2015 as a standalone ideograph with sequence num. 02833, source UTC-01039, and decomposition 日石双) anywhere within the consolidated lists above; to further complicate things, no encoded or WS2015-submitted ideograph uses this as a component (other than UTC-01039 itself). This author invites further discussion on how to deal with this component (whether to add it to §3n or to §4 above, or anything else?)

Appendix A: Rationale for Inclusion of Z3-03

Z3-03 丸 is included as a separate component from U+4E38 丸. First, they are considered different components in the Taiwan and Hong Kong locale. For characters whose etymology is U+4E2E 丮 (to hold), the glyph shape Z3-03 with the dot protruding is used. For characters whose etymology is U+4E38 丸 (small round object), the T/H glyph shape of U+4E38 with the dot not protruding is used.

This is a similar case to the existing dis-unified radicals 2EC0 ⺾ CJK RADICAL GRASS THREE from 2EBE 艹 CJK RADICAL GRASS ONE, and 2EBC 月 CJK RADICAL MEAT from 2F49 月 KANGXI RADICAL MOON. The difference in the shape is reflective of a different etymology and hence "non-cognateness" of characters in the Taiwan and the Hong Kong locales.

Second, these normally unifiable differences in components could also reflect a difference in meaning in certain context for other locales, and not just as a simple typographical variation. For example, PRC selectively applies such typographical variations as a standardized difference for a range of non-cognate characters. The difference between U+2EC0(⺾) and U+2EBE(艹) is used to distinguish between the two characters U+830D 茍 and U+82DF 苟 in the PRC locale.

Thus, the inclusion of Z3-03 component is also meaningful and useful for IDS composition.

Final thoughts:

Although the current immediate goal of Chan & Lu's IRG N2225 is to possibly establish a new Working Document Series (WDS), this author believes that, should these new components proposed above prove to be very useful in optimizing the IDS sequencing of currently-encoded CJK ideographs and future working-set and urgently-needed character submissions, these components should not just become a brand-new WDS, but also be proposed to WG2 as a new collection of CJK-related characters to be encoded into both the 13 unassigned codepoints of the **CJK Radicals Supplement** block (U+2E80 – U+2EFF) and a brand-new block suggested by Chan's IRG N2204 to be named **CJK Radicals Extended-A**. This author believes that such block be renamed to **CJK Components Supplement**, because none of these components proposed above (save for the six components with RADICAL in their provisional names, that is, H3-01, S9-01, P3-01, and P4-01 thru P4-03) are used as radicals but rather as phonetic components (or elements of phonetic components); also, this author suggests using up the empty 16-codepoint range **U+2FE0 – U+2FEF** (between the **Kangxi Radicals** [U+2F00 – U+2FDF] & **Ideographic Description Characters** [U+2FF0 – U+2FFF] blocks) to encode the 16 most commonly-used components among those listed above.

So, this author suggests that the IRG Rapporteur propose to WG2 the six unencoded radicals H3-01, S9-01, P3-01, and P4-01 thru P4-03 above to be encoded into any of the 13 free codepoints of **CJK Radicals Supplement** and then, depending on the usefulness of the remaining unencoded components in IRG's current and future IDS work, revise the consolidated lists above and propose to WG2 the creation of two new blocks: **CJK Components** (U+2FE0 – U+2FEF), where the 16 most frequently encountered of all components proposed above can be allocated, and **CJK Components Supplement** (it could be allotted into either the SMP, the SIP, or possibly the TIP, because there is no room left in the BMP for a new CJK-related block other than the **CJK Components** block suggested by this author), where the remaining components accepted from the lists above can be encoded. (It is worth noting that some of the components listed above have been also submitted to IRG Working Set 2015 as standalone ideographs, so whether they are considered for "duplicate" encoding as components or not will depend on how early they are encoded as ideographs into the future **CJK Unified Ideographs Extended-G** block.)

Since it has been previously proposed by the Unicode Consortium in IRG N2117 to create an IDS property into the Unihan Database (this author suggests making such property an informational one), the components proposed above, if accepted for encoding into the UCS, could be used to help populate this recently-proposed Unihan property. These components could also be incorporated into the U-source database (USourceData.txt) established as part of Unicode Standard Annex #45 (UAX45), which currently uses the "？" placeholder for indescribable, non-decomposable unencoded components.