IVS vs UCS

安岡孝一*

1　はじめに

IVS†がWindows 8やMac OS Xで大々的にサポートされた結果、これまで「文字コードのレベルでは見分けることができない」とされてきた軽微な字体差が、何となく交換可能になってきた気がする。すなわち「葛」と「葛」の差が、<845B E0100>と<845B E0101>というIVSによって、表現し分けられるような気がしてきたのである。同様に、「喝」と「喝」の差は、<559D E0100>と<559D E0101>で、「謁」と「謁」の差は、<8B01 E0101>と<8B01 E0100>で、それぞれ表現し分ければいい。ある意味、便利な時代になったと言えるだろう。

2　渇と渇の問題

ところが、そのようなIVSによる字体差表現から、ポツンと取り残されてしまった部分がある。いわゆる原規格分離‡に関わる部分で、例を挙げるなら「渇」と「渇」だ。「渇」と「渇」の差は、IVSではなく、従来どおり6E07と6E34というUCSレベルの差で表現される。あえてIVSで表現したとしても、<6E07 E0100>と<6E34 E0100>であり、IVSで表現し分けるわけではないのだ。同様に「掲」と「掲」の差も、63B2と63EDというUCSレベルの差で表現される。

渇渇 掲掲

<6E07 E0100> <6E34 E0100>　<63B2 E0100> <63ED E0100>

偈偈 喝喝 喝喝 愒愒

<5048 E0102> <5048 E0101>　<559D E0100> <559D E0101>　<559D E0102> <559D E0103>　<6112 E0102> <6112 E0101>

曷曷 歇歇 碣碣 竭竭

<66F7 E0102> <66F7 E0101>　<6B47 E0102> <6B47 E0101>　<78A3 E0102> <78A3 E0101>　<7AED E0102> <7AED E0101>

---

*京都大学人文科学研究所附属東アジア人文情報学研究センター

†Charts for the Unicode Ideographic Variation Database, Unicode Inc. (March 2, 2012)

‡ISO/IEC 10646 Annex S.3

羯羯 臈臈 藹藹 葛葛
<7FAF E0103> <7FAF E0101> <81C8 E0103> <81C8 E0101> <FA1F E0102> <FA1F E0101> <845B E0100> <845B E0101>

葛葛 葛葛 蝎蝎 蠍蠍
<845B E0102> <845B E0103> <845B E0107> <845B E0108> <874E E0102> <874E E0101> <880D E0103> <880D E0101>

褐褐 褐褐 謁謁 謁謁
<8910 E0100> <8910 E0101> <8910 E0102> <8910 E0103> <8B01 E0101> <8B01 E0100> <8B01 E0102> <8B01 E0103>

遏遏 遏遏 靄靄 鞨鞨
<904F E0103> <904F E0102> <904F E0104> <904F E0101> <9744 E0101> <9744 E0102> <97A8 E0102> <97A8 E0101>

この事実は、IVSを用いた漢字システムを設計・開発する際に、途方もないヤヤコシサを引き起こす。常識的に考えれば、「謁」と「謁」を検索上で同一視するシステムでは、「渇」と「渇」も同一視すべきだろう。逆に、「渇」と「渇」を分け隔てて処理するような局面では、「謁」と「謁」も分け隔てて処理すべきだろう。ところが、そう言うのは簡単だが、実現は途方もなく困難だ。片やIVSレベルの差で、片やUCSレベルの差なのだ。

3　飲と飲の問題

同様の問題を内包しているのが、「飲」と「飲」だ。食へんの新旧は、ほぼ全ての漢字においてIVSの差として表現されているが、「飲」と「飲」だけは、98F2と98EEというUCSレベルの差になっているのだ。

飲飲
<98F2 E0100> <98EE E0100>

飢飢 飣飣 飢飢 飢飢
<29682 E0101><29682 E0100> <29685 E0101><29685 E0100> <98E2 E0100> <98E2 E0101> <98E2 E0102> <98E2 E0104>

飣飣 飣飣 飦飦 飦飦 飣飣 飣飣 飣飣 飣飣
<98E3 E0102> <98E3 E0101> <4B22 E0101> <4B22 E0100> <98E6 E0102> <98E6 E0101> <29688 E0101><29688 E0100>

飣飣 飣飣 飩飩 飩飩 飪飪 飪飪 飪飪 飪飪
<29695 E0101><29695 E0100> <29696 E0101><29696 E0100> <98E9 E0102> <98E9 E0101> <98EA E0102> <98EA E0101>

飫飫 飫飫 飫飫 飫飫 飯飯 飯飯 飯飯 飯飯
<98EB E0101> <98EB E0100> <98EB E0103> <98EB E0102> <98ED E0102> <98ED E0101> <98EF E0101> <98EF E0100>

飯飯 飯飯 飠飠 飠飠 飴飴 飴飴 飴飴 飴飴
<98EF E0103> <98EF E0104> <4B2A E0101> <4B2A E0100> <4B2E E0101> <4B2E E0100> <98F4 E0100> <98F4 E0101>

飴飴 飴飴 飴飴 飴飴 飴飴 飴飴 飴飴 飴飴
<98F4 E0102> <98F4 E0103> <98F5 E0100> <98F5 E0101> <98F6 E0102> <98F6 E0101> <98FB E0100> <98FB E0101>

飼飼 飼飼 飼飼 飼飼 飾飾 飾飾 飾飾 飾飾
<98FC E0100> <98FC E0101> <98FC E0102> <98FC E0103> <98FD E0103> <98FD E0104> <98FE E0100> <98FE E0101>

飾飾 飾飾 飾飾 飾飾 飾飾 飾飾 飾飾 飾飾
<98FE E0102> <98FE E0103> <4B30 E0101> <4B30 E0100> <4B33 E0101> <4B33 E0100> <4B34 E0101> <4B34 E0100>

飴飴 飴飴 飴飴 飴飴 飴飴 飴飴 飴飴 飴飴
<296A9 E0101><296A9 E0100> <296B9 E0101><296B9 E0100> <296C6 E0101><296C6 E0100> <9901 E0101> <9901 E0100>

蝕蝕 蝕蝕 蝕蝕 蝕蝕 餅餅 餅餅 餅餅 餅餅
<8755 E0100> <8755 E0101> <8755 E0102> <8755 E0103> <9905 E0100> <9905 E0101> <9905 E0102> <9905 E0103>

餉餉 餉餉 餉餉 餉餉 餌餌 餌餌 餌餌 餌餌
<9909 E0101> <9909 E0100> <9909 E0103> <9909 E0102> <990C E0100> <990C E0103> <990C E0102> <990C E0101>

餌餌 餪餪 餖餖 餒餒
<990C E0104> <990C E0105> <4B39 E0101> <4B39 E0100> <9911 E0102> <9911 E0101> <9912 E0103> <9912 E0101>

餓餓 餓餓 餔餔 餕餕
<9913 E0100> <9913 E0101> <9913 E0102> <9913 E0103> <9914 E0102> <9914 E0101> <9915 E0103> <9915 E0101>

餖餖 餗餗 餘餘 餝餝
<9916 E0102> <9916 E0101> <9917 E0102> <9917 E0101> <9918 E0102> <9918 E0101> <4B3C E0101> <4B3C E0100>

餖餖 餜餜 餽餽 餬餬
<296DE E0101><296DE E0100> <296E5 E0101><296E5 E0100> <4B42 E0101> <4B42 E0100> <296FA E0101><296FA E0100>

餼餼 餚餚 餛餛 餜餜
<29706 E0101><29706 E0100> <991A E0102> <991A E0101> <991B E0102> <991B E0101> <991C E0102> <991C E0101>

餤餤 餞餞 餞餞 餟餟
<4B43 E0101> <4B43 E0100> <991E E0102> <991E E0101> <4B3B E0102> <4B3B E0101> <991F E0102> <991F E0101>

餅餅 餡餡 餡餡 餡餡
<9920 E0102> <9920 E0101> <9921 E0101> <9921 E0100> <9921 E0103> <9921 E0102> <9922 E0102> <9922 E0101>

餓餓 餤餤 餦餦 餧餧
<9923 E0100> <9923 E0101> <9924 E0102> <9924 E0101> <9926 E0102> <9926 E0101> <9927 E0102> <9927 E0101>

館館 館館 餩餩 餬餬
<9928 E0100> <9928 E0101> <9928 E0102> <9928 E0103> <9929 E0100> <9929 E0101> <4B45 E0101> <4B45 E0100>

餬餬 餬餬 餬餬 餬餬
<4B47 E0101> <4B47 E0100> <2970B E0101><2970B E0100> <2970F E0101><2970F E0100> <29717 E0101><29717 E0100>

饒饒 <29719 E0101><29719 E0102>
餕餕 <2972F E0101><2972F E0100>
餡餡 <29734 E0101><29734 E0100>
饌饌 <29739 E0101><29739 E0100>

餜餜 <2973F E0101><2973F E0100>
餪餪 <992A E0100> <992A E0101>
餫餫 <992B E0102> <992B E0101>
餬餬 <992C E0102> <992C E0101>

餭餭 <992D E0100> <992D E0101>
餰餰 <9930 E0100> <9930 E0101>
餱餱 <9931 E0102> <9931 E0101>
餲餲 <9932 E0102> <9932 E0101>

餳餳 <9933 E0102> <9933 E0101>
餴餴 <9934 E0102> <9934 E0103>
餴餴 <4B49 E0101> <4B49 E0100>
餵餵 <4B4B E0101> <4B4B E0100>

餶餶 <9936 E0100> <9936 E0101>
餹餹 <9939 E0103> <9939 E0102>
餻餻 <993B E0102> <993B E0101>
餻餻 <4B50 E0101> <4B50 E0100>

餼餼 <4B51 E0101> <4B51 E0100>
餼餼 <993C E0102> <993C E0101>
餽餽 <993D E0102> <993D E0101>
餾餾 <993E E0102> <993E E0101>

餿餿 <993F E0101> <993F E0100>
餡餡 <9940 E0102> <9940 E0101>
餿餿 <4B52 E0101> <4B52 E0100>
饁饁 <9941 E0102> <9941 E0101>

饂饂 <9942 E0102> <9942 E0103>
饂饂 <9942 E0104> <9942 E0101>
饁饁 <2975D E0101><2975D E0100>
饂饂 <29780 E0101><29780 E0100>

饈饈 <29783 E0101><29783 E0100>
餰餰 <2978F E0101><2978F E0100>
饉饉 <9944 E0100> <9944 E0101>
饉饉 <9944 E0100> <9944 E0101>

饅饅 <9945 E0101> <9945 E0100>
饅饅 <9945 E0105> <9945 E0103>
饆饆 <9946 E0102> <9946 E0101>
饇饇 <9947 E0102> <9947 E0101>

饈	饈	饉	饉	饑	饑	饋	饋
<9948 E0102>	<9948 E0101>	<9949 E0102>	<9949 E0101>	<994A E0101>	<994A E0100>	<994B E0101>	<994B E0100>
饋	饋	饌	饌	饒	饒	饎	饎
<994B E0103>	<994B E0102>	<994C E0102>	<994C E0101>	<994D E0102>	<994D E0101>	<994E E0102>	<994E E0101>
饐	饐	饑	饑	饒	饒	饑	饑
<9950 E0102>	<9950 E0101>	<9951 E0102>	<9951 E0101>	<9952 E0102>	<9952 E0101>	<9956 E0100>	<9956 E0101>
饌	饌	饑	饑	饑	饑	饑	饑
<29791 E0101>	<29791 E0100>	<297A1 E0101>	<297A1 E0100>	<297A5 E0101>	<297A5 E0100>	<297A7 E0101>	<297A7 E0100>
饌	饌	饑	饑	饑	饑	饑	饑
<297AB E0101>	<297AB E0100>	<297AD E0101>	<297AD E0100>	<297B7 E0101>	<297B7 E0100>	<297C4 E0101>	<297C4 E0100>
饑	饑	饗	饗	饗	饗	饗	饗
<297CB E0101>	<297CB E0100>	<9958 E0102>	<9958 E0101>	<9959 E0102>	<9959 E0101>	<995B E0103>	<995B E0102>
饑	饑	饗	饗	饗	饗	饗	饗
<4B61 E0101>	<4B61 E0100>	<995D E0100>	<995D E0101>	<297F1 E0101>	<297F1 E0100>	<995E E0102>	<995E E0101>
饗	饗	饗	饗	饗	饗	饗	饗
<995F E0102>	<995F E0101>	<9960 E0102>	<9960 E0101>	<9961 E0100>	<9961 E0101>	<4B63 E0101>	<4B63 E0100>
饗	饗	饗	饗	饗	饗		
<4B68 E0101>	<4B68 E0100>	<4B69 E0101>	<4B69 E0100>	<297FD E0100>	<297FD E0101>		

これは、JIS X 0208において、どういうわけか「飲」と「飲」に別々の区点番号が付与されていたことに、その原因がある。もともと食へんの新旧はUCSでは統合されることになっていたのに、原規格分離を日本が主張したために、UCSにおいても「飲」と「飲」だけが別々になってしまったのだ。不幸な歴史の結果だが、

まあ、98F2と98EEだけを特別に処理するしかないだろう。

4　研と研の問題

原規格分離は、日本の規格にだけ適用されたわけではない。たとえば「研」と「研」は、台湾のCNS 11643で別々の文字コードが付与されていたために、UCSにおいても7814と784Fという別コードとなってしまった。

刑刑 妍妍 研研 笄笄

<5211 E0100> <206EC E0100> <598D E0100> <59F8 E0100> <7814 E0100> <784F E0100> <7B04 E0100> <7B53 E0100>

邢邢

<90A2 E0100> <90C9 E0100>

呀呀 岍岍 形形 形形

<549E E0101> <549E E0100> <5C8D E0100> <5C8D E0101> <5F62 E0100> <5F62 E0101> <5F62 E0102> <5F62 E0103>

枅枅 汧汧 瓶瓶 眀眀

<6785 E0101> <6785 E0102> <6C67 E0102> <6C67 E0101> <3F1B E0100> <3F1B E0101> <4018 E0100> <4018 E0101>

秅秅 缾缾 袈袈 荊荊

<412F E0100> <412F E0101> <26228 E0100><26228 E0101> <27667 E0100><27667 E0101> <834A E0101> <834A E0104>

荊荊 荊荊 荊荊 蚈蚈

<834A E0103> <834A E0102> <8346 E0101> <8346 E0102> <8346 E0103> <8346 E0104> <8688 E0102> <8688 E0101>

貅貅 趼趼 鳽鳽

<27CA8 E0100><27CA8 E0101> <8DBC E0102> <8DBC E0101> <9CFD E0100> <9CFD E0101>

ただし、CNS 11643では「岍」と「岍」にも、別々の文字コードが付与されている。ところが、これらはUCSレベルでは統合されていて、<5C8D E0100>と<5C8D

E0101>というIVSレベルで見分けなければならない。原規格分離が必ずしも徹底されているわけではない、ということである。

5　温と溫の問題

同様に「温」と「溫」は、CNS 11643で別々の文字コードが付与されていたために、UCSにおいても6E29と6EABという別コードとなってしまった。あるいは「鰛」と「鰮」は、JIS X 0208で別々の区点番号が割り当てられていたために、UCSにおいても9C1Bと9C2Eという別コードとなってしまった。

愠	慍	昷	囩	温	溫	蒀	蒀
<6120 E0100>	<614D E0100>	<6637 E0100>	<25055 E0100>	<6E29 E0100>	<6EAB E0100>	<8480 E0100>	<8495 E0101>
蒀	蒀	蘊	蘊	蘊	蘊	蘊	蘊
<8480 E0101>	<8495 E0102>	<85F4 E0100>	<860A E0100>	<85F4 E0101>	<860A E0101>	<85F4 E0102>	<860A E0102>
鰛	鰮						
<9C1B E0100>	<9C2E E0100>						

嗢	嗢	暍	暍	瑥	瑥	瘟	瘟
<55E2 E0102>	<55E2 E0101>	<3B08 E0100>	<3B08 E0101>	<7465 E0101>	<7465 E0100>	<761F E0102>	<761F E0101>
蝹	蝹	褞	褞	韞	韞	韞	韞
<8779 E0101>	<8779 E0100>	<891E E0102>	<891E E0101>	<97DE E0101>	<97DE E0100>	<97DE E0104>	<97DE E0103>
餲	餲	餲	餲	馧	馧		
<9942 E0102>	<9942 E0104>	<9942 E0103>	<9942 E0101>	<99A7 E0101>	<99A7 E0100>		

ただし、CNS 11643では「暍」と「暍」にも、別々の文字コードが付与されている。ところが、これらはUCSレベルでは統合されていて、<3B08 E0100>と<3B08 E0101>というIVSレベルで見分けなければならない。原規格分離が必ずしも徹底されているわけではない、ということである。

6　横と横の問題

「横」と「横」は、UCSレベルにおいて6A2Aと6A6Bという別々の文字コードとなってしまっている。一方、「黄」と「黄」を部品とするその他の文字は、いずれもIVDレベルでの差となっている。「横」と「横」の6A2Aと6A6Bを、特別に処理するしかないだろう。

横	横	黄	黄	黄	黄
<6A2A E0100>	<6A6B E0100>	<9EC4 E0100>	<9EC3 E0100>	<9EC4 E0101>	<9EC3 E0101>

僥	僥	勵	勵	嚝	嚝	壙	壙
<50D9 E0103>	<50D9 E0101>	<208B9 E0102>	<208B9 E0101>	<569D E0102>	<569D E0101>	<58D9 E0102>	<58D9 E0101>
廣	廣	廣	廣	撗	撗	擴	擴
<5EE3 E0101>	<5EE3 E0100>	<5EE3 E0105>	<5EE3 E0103>	<6497 E0101>	<6497 E0100>	<64F4 E0102>	<64F4 E0101>
曠	曠	櫎	櫎	潢	潢	濂	濂
<66E0 E0102>	<66E0 E0101>	<6ACE E0101>	<6ACE E0100>	<6F62 E0103>	<6F62 E0101>	<7007 E0102>	<7007 E0101>
瀇	瀇	熿	熿	黶	黶	璜	璜
<3D87 E0101>	<3D87 E0100>	<71BF E0102>	<71BF E0101>	<244BC E0101>	<244BC E0100>	<749C E0102>	<749C E0101>
曠	曠	礦	礦	簧	簧	纊	纊
<77CC E0102>	<77CC E0100>	<7926 E0102>	<7926 E0101>	<7C27 E0102>	<7C27 E0101>	<7E8A E0102>	<7E8A E0101>
糤	糤	贕	贕	鐄	鐄	鑛	鑛
<42A3 E0101>	<42A3 E0100>	<27E3C E0100>	<27E3C E0101>	<9404 E0102>	<9404 E0101>	<945B E0102>	<945B E0101>
騿	騿	黻	黻	黌	黌		
<4BB2 E0101>	<4BB2 E0100>	<9EC6 E0104>	<9EC6 E0101>	<9ECC E0103>	<9ECC E0102>		

7　渉と渉の問題

「渉」と「渉」は、UCSレベルで6E09と6D89という別々の文字コードが付与されている。一方、「歩」と「歩」を部品とするその他の文字は、いずれもIVDレベルでの差となっている。「渉」と「渉」の6E09と6D89を、特別に処理するしかないだろう。

歩歩 渉渉

<6B69 E0100> <6B65 E0100>　<6E09 E0100> <6D89 E0100>

徙徙 捗捗 捗捗 頻頻

<5F8F E0102> <5F8F E0101>　<6357 E0100> <6357 E0101>　<6357 E0102> <6357 E0103>　<983B E0101> <983B E0100>

頻頻 陟陟

<983B E0102> <983B E0103>　<965F E0102> <965F E0101>

8　説と說の問題

「説」と「說」は、UCSレベルで8AACと8AAAという別々の文字コードが付与されている。一方、「唳」と「唳」は、UCSレベルでは54FEに統合されていて、見分けるためにはIVSが必要である。

悦悦 稅稅 脫脫 說說

<60A6 E0100> <6085 E0100>　<7A0E E0100> <7A05 E0100>　<8131 E0100> <812B E0100>　<8AAC E0100> <8AAA E0100>

鋭鋭 閲閲

<92ED E0100> <92B3 E0100>　<95B2 E0100> <95B1 E0100>

倪倪 唳唳 埕埕 娩娩

<4FBB E0103> <4FBB E0101>　<54FE E0102> <54FE E0101>　<3642 E0101> <3642 E0100>　<5A27 E0102> <5A27 E0101>

帨	帨	裞	裞	覷	覷
<5E28 E0102>	<5E28 E0101>	<7971 E0101>	<7971 E0102>	<271D3 E0102>	<271D3 E0101>

「兑」と「兌」を部品とする文字は、UCSレベルで分離されているものと、IVSレベルで分離されているものが、拮抗しているように思える。どう考えても、ケースバイケースで処理するしかなく、かなりヤヤコシイことこの上ない。

9　錬と鍊の問題

「棟」と「楝」は、UCSレベルで68DFと695Dという別々の文字コードが付与されている。一方、「練」と「練」は、UCSレベルでは7DF4に統合されていて、見分けるためにはIVSが必要である。つまり、部品としての「東」と「柬」は、文字によって扱いが違う。

というのも、そもそも「東」と「柬」を部品とする文字どうしは、時には別字となったり、時には異体字となったりする。別字の場合には、かなりの確率でUCSレベルでの分離がおこなわれてしまう。異体字の場合には、IVSレベルでの分離になる場合が多いが、必ずしもそうなるとは限らない。たとえ異体字であっても、「錬」と「鍊」のようにUCSレベルで分離されてしまう場合もあるからだ。ケースバイケースで処理しなければならない、という点で、非常にヤヤコシイ。

東	柬	棟	楝	涷	湅	諫	諫
<6771 E0100>	<67EC E0100>	<68DF E0100>	<695D E0100>	<6DB7 E0100>	<6E45 E0100>	<8ACC E0100>	<8AEB E0100>
錬	鍊	鰊	鰊	鶇	鶇		
<932C E0100>	<934A E0100>	<9BDF E0100>	<9C0A E0100>	<9D87 E0100>	<9DAB E0100>		

悚	悚	攔	攔	欄	欄	欄	欄
<226D4 E0100>	<226D4 E0101>	<6514 E0102>	<6514 E0101>	<6B04 E0101>	<6B04 E0100>	<6B04 E0102>	<6B04 E0103>
欗	欗	瀾	瀾	灡	灡	煉	煉
<6B17 E0103>	<6B17 E0101>	<703E E0102>	<703E E0101>	<7061 E0102>	<7061 E0100>	<7149 E0101>	<7149 E0100>

煉煉 爛爛 爛爛 糠糠
<7149 E0102> <7149 E0103> <721B E0102> <721B E0101> <7224 E0103> <7224 E0102> <25EC2 E0102><25EC2 E0101>

練練 練練 蘭蘭 蘭蘭
<7DF4 E0101> <7DF4 E0100> <7DF4 E0102> <7DF4 E0103> <862D E0100> <862D E0101> <862D E0102> <862D E0104>

蘭蘭
<862D E0105> <862D E0103>

10　鄉と鄉の問題

「鄉」と「鄉」の関係は、正直わけのわからないことになっている。<9115 E0101>と<9115 E0102>というIVSレベルの差としても表せるし、<9115 E0100>と<90F7 E0101>というUCSレベルの差でもあるのだ。これは、Adobe-Japan1とHanyo-Denshiとの間で、異なる立場を採用してしまったことに起因する。すなわち、Adobe-Japan1は<9115 E0100>と<90F7 E0101>というUCSレベルの差を採用したのに対し、Hanyo-Denshiは<9115 E0101>と<9115 E0102>というIVSレベルの差を採用しているのだ。その結果「IVS vs UCS」とも言える対立が発生していて、どちらを使うかはユーザ次第という、かなり悲しい事態となっている。

既既 鄉鄉
<65E2 E0103> <65E3 E0102> <9115 E0100> <90F7 E0101>

卽卽 卿卿 卿卿 厩厩
<537D E0102> <537D E0101> <537F E0102> <537F E0100> <537F E0104> <537F E0106> <53A9 E0108> <53A9 E0109>

廏廏 廏廏 概概 節節
<5ECF E0101> <5ECF E0100> <5ECF E0103> <5ECF E0102> <69EA E0103> <69EA E0104> <7BC0 E0101> <7BC0 E0102>

節節 鄉鄉

<7BC0 E0105> <7BC0 E0104> <9115 E0101> <9115 E0102>

11　併と併の問題

「併」と「併」の関係も、かなりわけがわからない。<4F75 E0100>と<4F75 E0101>というIVSレベルの差としても表せるし、<4F75 E0100>と<5002 E0102>というUCSレベルの差でも表せるのだ。まあ、もともとの部品である「并」と「幷」が、5E76と5E77というUCSレベルの差なので、「併」と「併」もUCSの差として表した方が、あるいはスジがいいのかもしれない。

併併 屏屏 并幷 瓶瓶

<4F75 E0100> <5002 E0102> <5C4F E0100> <5C5B E0100> <5E76 E0100> <5E77 E0100> <74F6 E0100> <7501 E0100>

胼胼 逆逆 逆逆 逆逆

<80FC E0100> <8141 E0100> <8FF8 E0100> <902C E0100> <8FF8 E0101> <902C E0101> <8FF8 E0102> <902C E0102>

餅餅 餅餅 餅餅

<9905 E0102> <9920 E0102> <9905 E0101> <9920 E0100> <9905 E0103> <9920 E0101>

併併 塀塀 塀塀 拼拼

<4F75 E0100> <4F75 E0101> <5840 E0100> <5840 E0101> <5840 E0102> <5840 E0103> <62FC E0101> <62FC E0102>

摏摏 栟栟 栟栟 洴洴

<6452 E0102> <6452 E0101> <681F E0100> <681F E0101> <681F E0102> <681F E0103> <6D34 E0102> <6D34 E0101>

竝竝 秤秤 艏艏 絣絣

<7AEE E0100> <7AEE E0101> <257A9 E0101><257A9 E0102> <446B E0100> <446B E0101> <7D63 E0100> <7D63 E0101>

12　杞と杞の問題

「杞」と「杞」は、悲しい歴史を背負っている。これは元々、Adobe-Japan1が、<675E E0100>と<675E E0101>というIVSレベルの差だとしていたものだ。しかし、「杞」が233CCに既収録であるとの指摘を受けて、<233CC E0100>にも同じ「杞」を追加登録したものである。したがって、Adobe-Japan1としては、現在は<675E E0100>と<233CC E0100>というUCSレベルの差を使ってほしいのだが、歴史的経緯もあって、<675E E0100>と<675E E0101>というIVSレベルの差でもかまわない、ということにしているのである。今後も、こういう「IVS vs UCS」的な追加がおこなわれる可能性は十分にあって、正直なところ気が重い。

熙熙 熙熙 熙熙 玘玘
<7199 E0102> <7199 E0101> <242EE E0101><242EE E0100> <242EE E0103><242EE E0102> <7398 E0101> <7398 E0102>

祀祀 祀祀 紀紀 紀紀
<7940 E0103> <7940 E0105> <7940 E0104> <7940 E0102> <7D00 E0101> <7D00 E0104> <7D00 E0106> <7D00 E0105>

記記 配配
<8A18 E0101> <8A18 E0102> <914D E0102> <914D E0103>

13　嘕と嘕の問題

「間」と「閒」は、UCSでは別コードの9593と9592となっている。「間」と「閒」を部品として含む文字も、それぞれUCSレベルで別コードとなっている。ところが、「嘕」と「嘕」だけは、<35F4 E0101>と<35F4 E0102>というIVSレベルでの差となっている。実は、JIS X 0213の「嘕」をUCSに追加提案する際、うっかり35F4に統合してしまったのだが、今にして思うと、あれはミスだったのではないかという気がしてならない。

僩僩 澗澗 燗燗 癇癇
<50F4 E0100> <50E9 E0100> <6F97 E0100> <3D4E E0100> <243D0 E0100><71D7 E0100> <764E E0100> <7647 E0100>

瞷瞷 蕑蕑 蕑蕑 間閒
<77B7 E0100> <77AF E0100> <8573 E0100> <8551 E0102> <8573 E0101> <8551 E0103> <9593 E0100> <9592 E0100>

鬩鬩
<9B1D E0100> <9B1C E0100>

嘕嘕
<35F4 E0101> <35F4 E0102>

14　おわりに

同一の字体差が、IVSレベルでの分離となったり、UCSレベルでの分離となったりする場合を、いくつかの例について見てきた。UCSとIVSが、それぞれ行き当たりばったりに拡張されてきた結果、こういうヤヤコシイことになっているのである。ただ、こうなってしまった以上、もはや、これらをスッキリさせるのは難しく、あるがままにケースバイケースで処理するしかない、という点が、何とも気が重い。

なお、本稿では、あくまで現時点で筆者の目にとまった文字のうち、IVSが定義されているものだけを示しており、全ての「IVS vs UCS」な関係を網羅しているわけではない。今後も作業を続けたいとは思うが、そもそも網羅不可能ではないかと思えるほどの文字を扱わねばならず、たとえば「IVSに対するIDS記述」のような、何かうまいやり方を、今後は見つける必要があるように思う。