

# Adobe-Japan1 IVD Collection: Current Status & Future Directions

Dr. Ken Lunde

Senior Computer Scientist

Adobe Systems Incorporated

[lunde@adobe.com](mailto:lunde@adobe.com)



# What Is An Ideographic Variation Sequence?

- Base Character + Variation Selector = Glyph
  - Base Character (BC)—Any “CJK Unified Ideograph”
  - Variation Selector (VS)—U+E0100 through U+E01EF (240 VSes in total)
    - Designated VS17 through VS256
    - *Note that U+FE00 through U+FE0F (VS1 through VS16) are not for IVS use!*
- An IVS resolves to a glyph
- An IVS is registered and unique
  - Ideographic Variation Database (IVD) collections undergo a registration process
- The power, safety, and reliability of “plain text”
  - The ability to survive or endure in more environments
- See: <http://www.unicode.org/ivd/>

# “Adobe-Japan1” IVD Collection: Current Status

- Based on the Adobe-Japan1-6 Character Collection
  - 23,058 glyphs in total—CIDs (Character IDs) 0 through 23057
  - 14,665 of these glyphs are classified as ideographs
- “Adobe-Japan1” IVD Collection registration history and summary
  - PRI 98 issued in 12/2006—*First Draft*
  - PRI 108 issued in 08/2007—*Second Draft*
  - Final form declared on 12/14/2007—14,647 registered IVSes covering 14,645 ideographs
  - Two duplicate IVSes—*two or more different IVSes mapping to the same source (glyph)*
- 20 ideographs remain without registered IVSes
  - One was in Extension C
  - The remaining 19 were submitted as new CJK Unified Ideographs via IRG N1373 (IRG 29)
  - These 19 ideographs were added to the UNC set, which later became Extension D
    - 17 of these 20 ideographs are now included in Extension D

# “Adobe-Japan1” IVD Collection: IVSes To Be Registered

- Unified with existing CJK Unified Ideograph (IRG N1677; IRG 34)

5DE5 E0101; Adobe-Japan1; CID+13763

互 ↔ 工

- Found in Extension B

233CC E0100; Adobe-Japan1; CID+14140—*duplicate*

- Unicode 5.2

9FC4 E0100; Adobe-Japan1; CID+14089—*duplicate*

9FC6 E0100; Adobe-Japan1; CID+14168—*duplicate*

- Extension C

2A9E6 E0100; Adobe-Japan1; CID+14145

# “Adobe-Japan1” IVD Collection: IVSes To Be Registered (cont’d)

- Extension D

2B746 E0100; Adobe-Japan1; CID+13780—*duplicate*

2B751 E0100; Adobe-Japan1; CID+13866—*duplicate*

2B753 E0100; Adobe-Japan1; CID+20088

2B75A E0100; Adobe-Japan1; CID+20096

2B75C E0100; Adobe-Japan1; CID+20097

2B765 E0100; Adobe-Japan1; CID+20247

2B776 E0100; Adobe-Japan1; CID+20114—*duplicate*

2B777 E0100; Adobe-Japan1; CID+13782

2B77C E0100; Adobe-Japan1; CID+20125

2B782 E0100; Adobe-Japan1; CID+20141

2B789 E0100; Adobe-Japan1; CID+14064—*duplicate*

2B78B E0100; Adobe-Japan1; CID+20149

# “Adobe-Japan1” IVD Collection: IVSes To Be Registered (cont’d)

2B78E E0100; Adobe-Japan1; CID+13724—*duplicate*

2B794 E0100; Adobe-Japan1; CID+20153

2B7AC E0100; Adobe-Japan1; CID+20176

2B7AF E0100; Adobe-Japan1; CID+20180

2B7BD E0100; Adobe-Japan1; CID+14174

2B7C9 E0100; Adobe-Japan1; CID+20194

2B7CF E0100; Adobe-Japan1; CID+20201—*duplicate*

2B7D2 E0100; Adobe-Japan1; CID+20204

2B7D8 E0100; Adobe-Japan1; CID+13651—*duplicate*

2B7F0 E0100; Adobe-Japan1; CID+20240—*duplicate*

2B80D E0100; Adobe-Japan1; CID+20256

2B817 E0100; Adobe-Japan1; CID+20260

2B81A E0100; Adobe-Japan1; CID+14278

# “Adobe-Japan1” IVD Collection: IVSes To Be Registered (cont’d)

- To be included post Extension E (IRG N1677; IRG 34)

- CID+20156—variant of U+6DBC/CID+3979

涼 ↔ 涼

# “Adobe-Japan1” IVD Collection: Duplicate IVSes

- Existing duplicate IVSes (two)

9039 E0101; Adobe-Japan1; CID+13912 & 9054 E0101; Adobe-Japan1; CID+13912

29FCE E0100; Adobe-Japan1; CID+19071 & 29FD7 E0100; Adobe-Japan1; CID+19071

- Future duplicate IVSes—found in Extension B (one)

675E E0101; Adobe-Japan1; CID+14140 & 233CC E0100; *Adobe-Japan1; CID+14140*

- Future duplicate IVSes—Unicode 5.2 (two)

6881 E0101; Adobe-Japan1; CID+14089 & 9FC4 E0100; *Adobe-Japan1; CID+14089*

7953 E0101; Adobe-Japan1; CID+14168 & 9FC6 E0100; *Adobe-Japan1; CID+14168*



# “Adobe-Japan1” IVD Collection: Duplicate IVSes (cont’d)

- Future duplicate IVSes—Extension D (eight)

4ECA E0101; Adobe-Japan1; CID+13780 & 2B746 E0100; Adobe-Japan1; CID+13780

52E2 E0101; Adobe-Japan1; CID+13866 & 2B751 E0100; Adobe-Japan1; CID+13866

5EA7 E0101; Adobe-Japan1; CID+20114 & 2B776 E0100; Adobe-Japan1; CID+20114

687A E0101; Adobe-Japan1; CID+14064 & 2B789 E0100; Adobe-Japan1; CID+14064

83DF E0103; Adobe-Japan1; CID+20201 & 2B7CF E0100; Adobe-Japan1; CID+20201

885E E0101; Adobe-Japan1; CID+13651 & 2B7D8 E0100; Adobe-Japan1; CID+13651

943A E0101; Adobe-Japan1; CID+20240 & 2B7F0 E0100; Adobe-Japan1; CID+20240

2363A E0102; Adobe-Japan1; CID+13724 & 2B78E E0100; Adobe-Japan1; CID+13724

# The “Hanyo-Denshi” IVD Collection

- The Japan NB submitted their IVD collection for registration in late 03/2010
  - Designated PRI 167, with an expected closing date of 06/25/2010
  - 4,214 glyphs: 1,924 default forms plus 2,290 variant forms
  - My comments were officially submitted on 05/10/2010 as L2/10-196
- Relationship with the “Adobe-Japan1” IVD Collection
  - The 1,924 default forms unambiguously map directly to “Adobe-Japan1” IVSes
  - 636 of the 2,290 variant forms map to “Adobe-Japan1” IVSes
  - My mapping table was officially submitted on 06/03/2010
  - The two IVD collections are largely parallel, because...
    - Both IVD collections are intended for use in Japan
    - Both IVD collections include many of the same variant form

# IVS-Enabled Infrastructure

## ■ Fonts

- OpenType fonts can now include a Format 14 'cmap' subtable that maps UVSes to GIDs
  - Default versus non-default UVSes
  - Works with existing Format 4 or 12 'cmap' subtables for handling default UVSes
- Adobe Systems has developed fifty IVS-enabled OpenType Japanese fonts
  - Heisei, Kozuka Mincho/Gothic, and Ryo Text/Display/Gothic families
- Adobe Systems' AFDKO includes IVS-enabling font tools
- See: <http://www.adobe.com/devnet/opentype/afdko/>

## ■ OSes

- Mac OS X 10.6 and Windows 7

## ■ Applications

- Acrobat & Adobe Reader 9, InDesign CS4, and Flash Player 10

## ■ IVS-enabled Input Methods represent a missing infrastructure component

# IVD Collection Interaction

- UTS #37 now permits joint authority for IVSes through mutual consent
  - IVSes can be shared across multiple IVD collections
  - Involves two or more IVD collection registrants
  - Each IVD collection preserves its identity through its own sources/glyphs
  - Joint authority is not forced; requires mutual agreement
- Recommended for IVD collections that are largely parallel
- Reduces the number of duplicate IVSes that serve the same purpose
- See: <http://unicode.org/standard/reports/tr37/>

# Future IVD Collection Ideas

- CJK Compatibility Ideographs
  - All CJK Compatibility Ideographs are crippled due to Normalization
  - Joint authority for those already covered by existing IVD collections
    - “Adobe-Japan1” IVD Collection
  - 75 CJK Compatibility Ideographs correspond to JIS X 0213:2004 sources
    - 57 of these 75 CJK Compatibility Ideographs are included in the *Jinmei-yō Kanji* set (985)
- Pan-CJK Ideographs
  - Serves as the basis for developing Pan-CJK fonts
  - Allows Pan-CJK glyph distinctions to be preserved in “plain text”



**Adobe**