Genuine Han Unification

Dr. Ken Lunde | Senior Computer Scientist | Adobe Systems Incorporated



This presentation contains events that may happen...



This presentation contains events that *may* happen... ...not those that necessarily *will* happen

This presentation contains events that *may* happen... ...not those that necessarily *will* happen





Pop Quiz—Chinese or Japanese?



Pop Quiz—Chinese or Japanese?





























East Asian writing systems have changed over time, through reform

- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary

- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hanzi simplification of the 1950s

- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hànzì simplification of the 1950s— $_{
 m III}$ → 头

- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hànzì simplification of the 1950s— $i g \rightarrow$ 头
- Character form (glyph) preferences change over time

- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hànzì simplification of the 1950s— $_{
 m III}$ → 头
- Character form (glyph) preferences change over time
 - What was once deemed unacceptable is now the norm

- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hànzì simplification of the 1950s— $_{
 m III}$ → 头
- Character form (glyph) preferences change over time
 - What was once deemed unacceptable is now the norm
- Today's East Asian fonts fall into two categories

- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hànzì simplification of the 1950s— $i g \rightarrow$ 头
- Character form (glyph) preferences change over time
 - What was once deemed unacceptable is now the norm
- Today's East Asian fonts fall into two categories
 - Single-region fonts with a single glyph per CJK Unified Ideograph

- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hànzì simplification of the 1950s— $_{
 m III}$ → 头
- Character form (glyph) preferences change over time
 - What was once deemed unacceptable is now the norm
- Today's East Asian fonts fall into two categories
 - Single-region fonts with a single glyph per CJK Unified Ideograph
 - Multiple-region (aka, Pan-CJK) fonts with multiple glyphs per CJK Unified Ideograph



- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hànzì simplification of the 1950s— $_{
 m III}$ → 头
- Character form (glyph) preferences change over time
 - What was once deemed unacceptable is now the norm
- Today's East Asian fonts fall into two categories
 - Single-region fonts with a single glyph per CJK Unified Ideograph
 - Multiple-region (aka, Pan-CJK) fonts with multiple glyphs per CJK Unified Ideograph
 - Requires a lot of time and effort, and can lead to extrapolation



- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hànzì simplification of the 1950s— $_{
 m III}$ → 头
- Character form (glyph) preferences change over time
 - What was once deemed unacceptable is now the norm
- Today's East Asian fonts fall into two categories
 - Single-region fonts with a single glyph per CJK Unified Ideograph
 - Multiple-region (aka, Pan-CJK) fonts with multiple glyphs per CJK Unified Ideograph
 - Requires a lot of time and effort, and can lead to extrapolation
 - Some are useful only for a single region, but may need to be rendered for other regions



- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hànzì simplification of the 1950s— $i g \rightarrow$ 头
- Character form (glyph) preferences change over time
 - What was once deemed unacceptable is now the norm
- Today's East Asian fonts fall into two categories
 - Single-region fonts with a single glyph per CJK Unified Ideograph
 - Multiple-region (aka, Pan-CJK) fonts with multiple glyphs per CJK Unified Ideograph
 - Requires a lot of time and effort, and can lead to extrapolation
 - Some are useful only for a single region, but may need to be rendered for other regions
- East Asian writing systems will change in the future

- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hànzì simplification of the 1950s— $i g \rightarrow$ 头
- Character form (glyph) preferences change over time
 - What was once deemed unacceptable is now the norm
- Today's East Asian fonts fall into two categories
 - Single-region fonts with a single glyph per CJK Unified Ideograph
 - Multiple-region (aka, Pan-CJK) fonts with multiple glyphs per CJK Unified Ideograph
 - Requires a lot of time and effort, and can lead to extrapolation
 - Some are useful only for a single region, but may need to be rendered for other regions
- East Asian writing systems *will* change in the future
 - Can Unicode act as a catalyst for a single-glyph model that serves all regions?

- East Asian writing systems have changed over time, through reform
 - Reform is sometimes revolutionary—the hànzì simplification of the 1950s— $i g \rightarrow$ 头
- Character form (glyph) preferences change over time
 - What was once deemed unacceptable is now the norm
- Today's East Asian fonts fall into two categories
 - Single-region fonts with a single glyph per CJK Unified Ideograph
 - Multiple-region (aka, Pan-CJK) fonts with multiple glyphs per CJK Unified Ideograph
 - Requires a lot of time and effort, and can lead to extrapolation
 - Some are useful only for a single region, but may need to be rendered for other regions
- East Asian writing systems *will* change in the future
 - Can Unicode act as a catalyst for a single-glyph model that serves all regions?
 - Will greater cross-cultural communication play a role?

• "Unification" of the most important East Asian ("Han") character set standards

- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)

- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)
 - The end result is a single repertoire

- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)
 - The end result is a single repertoire: CJK Unified Ideographs

- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)
 - The end result is a single repertoire: CJK Unified Ideographs
- There are subtle—and sometimes not-so-subtle—region-specific forms

- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)
 - The end result is a single repertoire: CJK Unified Ideographs
- There are subtle—and sometimes not-so-subtle—region-specific forms
 - This is what is being unified

- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)
 - The end result is a single repertoire: CJK Unified Ideographs
- There are subtle—and sometimes not-so-subtle—region-specific forms
 - This is what is being unified
- Han Unification is based on a set of rules and principles

- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)
 - The end result is a single repertoire: CJK Unified Ideographs
- There are subtle—and sometimes not-so-subtle—region-specific forms
 - This is what is being unified
- Han Unification is based on a set of rules and principles
 - The notion of "Same Abstract Character" plays a significant role


- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)
 - The end result is a single repertoire: CJK Unified Ideographs
- There are subtle—and sometimes not-so-subtle—region-specific forms
 - This is what is being unified
- Han Unification is based on a set of rules and principles
 - The notion of "Same Abstract Character" plays a significant role
- Naturally, there are exceptions



- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)
 - The end result is a single repertoire: CJK Unified Ideographs
- There are subtle—and sometimes not-so-subtle—region-specific forms
 - This is what is being unified
- Han Unification is based on a set of rules and principles
 - The notion of "Same Abstract Character" plays a significant role
- Naturally, there are exceptions
 - Source Separation Rule



- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)
 - The end result is a single repertoire: CJK Unified Ideographs
- There are subtle—and sometimes not-so-subtle—region-specific forms
 - This is what is being unified
- Han Unification is based on a set of rules and principles
 - The notion of "Same Abstract Character" plays a significant role
- Naturally, there are exceptions
 - Source Separation Rule
 - Applicable only to the URO (*Unified Repertoire & Ordering*) in its original form

- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)
 - The end result is a single repertoire: CJK Unified Ideographs
- There are subtle—and sometimes not-so-subtle—region-specific forms
 - This is what is being unified
- Han Unification is based on a set of rules and principles
 - The notion of "Same Abstract Character" plays a significant role
- Naturally, there are exceptions
 - Source Separation Rule
 - Applicable only to the URO (*Unified Repertoire & Ordering*) in its original form
 - Simplified forms

- "Unification" of the most important East Asian ("Han") character set standards
 - Hànzì (Chinese), kanji (Japanese), and hanja (Korean)
 - The end result is a single repertoire: CJK Unified Ideographs
- There are subtle—and sometimes not-so-subtle—region-specific forms
 - This is what is being unified
- Han Unification is based on a set of rules and principles
 - The notion of "Same Abstract Character" plays a significant role
- Naturally, there are exceptions
 - Source Separation Rule
 - Applicable only to the URO (*Unified Repertoire & Ordering*) in its original form
 - Simplified forms
 - Some are unified (骨 vs 骨), but some are not (車 vs 车)

Block	Code Points	Characters	Version
URO	U+4E00-U+9FA5	20,902	1.1
CJK Compatibility Ideographs	U+FA0E, U+FA0F, U+FA11, U+FA13, U+FA14, U+FA1F, U+FA21, U+FA23, U+FA24, U+FA27–U+FA29	12	1.1

Block	Code Points	Characters	Version
URO	U+4E00-U+9FA5	20,902	1.1
CJK Compatibility Ideographs	U+FA0E, U+FA0F, U+FA11, U+FA13, U+FA14, U+FA1F, U+FA21, U+FA23, U+FA24, U+FA27–U+FA29	12	1.1
Extension A	U+3400-U+4DB5	6,582	3.0

Block	Code Points	Characters	Version
URO	U+4E00-U+9FA5	20,902	1.1
CJK Compatibility Ideographs	U+FA0E, U+FA0F, U+FA11, U+FA13, U+FA14, U+FA1F, U+FA21, U+FA23, U+FA24, U+FA27–U+FA29	12	1.1
Extension A	U+3400-U+4DB5	6,582	3.0
Extension B	U+20000-U+2A6D6	42,711	3.1

Block	Code Points	Characters	Version
URO	U+4E00-U+9FA5	20,902	1.1
CJK Compatibility Ideographs	U+FA0E, U+FA0F, U+FA11, U+FA13, U+FA14, U+FA1F, U+FA21, U+FA23, U+FA24, U+FA27–U+FA29	12	1.1
Extension A	U+3400-U+4DB5	6,582	3.0
Extension B	U+20000-U+2A6D6	42,711	3.1
URO—appended	U+9FA6-U+9FBB	22	4.1

Block	Code Points	Characters	Version
URO	U+4E00-U+9FA5	20,902	1.1
CJK Compatibility Ideographs	U+FA0E, U+FA0F, U+FA11, U+FA13, U+FA14, U+FA1F, U+FA21, U+FA23, U+FA24, U+FA27–U+FA29	12	1.1
Extension A	U+3400-U+4DB5	6,582	3.0
Extension B	U+20000-U+2A6D6	42,711	3.1
URO—appended	U+9FA6-U+9FBB	22	4.1
URO—appended	U+9FBC-U+9FC3	8	5.1

Block	Code Points	Characters	Version
URO	U+4E00-U+9FA5	20,902	1.1
CJK Compatibility Ideographs	U+FA0E, U+FA0F, U+FA11, U+FA13, U+FA14, U+FA1F, U+FA21, U+FA23, U+FA24, U+FA27–U+FA29	12	1.1
Extension A	U+3400-U+4DB5	6,582	3.0
Extension B	U+20000-U+2A6D6	42,711	3.1
URO—appended	U+9FA6-U+9FBB	22	4.1
URO—appended	U+9FBC-U+9FC3	8	5.1
URO—appended	U+9FC4–U+9FCB	8	5.2
Extension C	U+2A700-U+2B734	4,149	5.2

Block	Code Points	Characters	Version
URO	U+4E00-U+9FA5	20,902	1.1
CJK Compatibility Ideographs	U+FA0E, U+FA0F, U+FA11, U+FA13, U+FA14, U+FA1F, U+FA21, U+FA23, U+FA24, U+FA27–U+FA29	12	1.1
Extension A	U+3400-U+4DB5	6,582	3.0
Extension B	U+20000-U+2A6D6	42,711	3.1
URO—appended	U+9FA6-U+9FBB	22	4.1
URO—appended	U+9FBC-U+9FC3	8	5.1
URO—appended	U+9FC4–U+9FCB	8	5.2
Extension C	U+2A700-U+2B734	4,149	5.2
Extension D	U+2B740-U+2B81D	222	6.0

Block	Code Points	Characters	Version
URO	U+4E00-U+9FA5	20,902	1.1
CJK Compatibility Ideographs	U+FA0E, U+FA0F, U+FA11, U+FA13, U+FA14, U+FA1F, U+FA21, U+FA23, U+FA24, U+FA27–U+FA29	12	1.1
Extension A	U+3400-U+4DB5	6,582	3.0
Extension B	U+20000-U+2A6D6	42,711	3.1
URO—appended	U+9FA6-U+9FBB	22	4.1
URO—appended	U+9FBC-U+9FC3	8	5.1
URO—appended	U+9FC4–U+9FCB	8	5.2
Extension C	U+2A700-U+2B734	4,149	5.2
Extension D	U+2B740-U+2B81D	222	6.0
URO—appended	U+9FCC	1	6.1

Minor glyph differences do not prevent mutual intelligibility

Minor glyph differences do not prevent mutual intelligibility

写真 vs 写真

- Minor glyph differences do not prevent mutual intelligibility
 - 写真 vs 写真 京都 vs 京都

Minor glyph differences do not prevent mutual intelligibility

写真 vs 写真 京都 vs 京都 骨 vs 骨

Minor glyph differences do not prevent mutual intelligibility

写真 vs 写真 京都 vs 京都 骨 vs 骨

Simplified Chinese resulted in greater orthographic complexity

Minor glyph differences do not prevent mutual intelligibility

写真 vs 写真 京都 vs 京都 骨 vs 骨

Simplified Chinese resulted in greater orthographic complexity

漢 vs 汉

Minor glyph differences do not prevent mutual intelligibility

写真 vs 写真 京都 vs 京都 骨 vs 骨

Simplified Chinese resulted in greater orthographic complexity

漢 vs 汉 頭 vs 头

Minor glyph differences do not prevent mutual intelligibility

写真 vs 写真 京都 vs 京都 骨 vs 骨

Simplified Chinese resulted in greater orthographic complexity

漢vs汉 頭vs头 樂vs乐

Minor glyph differences do not prevent mutual intelligibility

写真 vs 写真 京都 vs 京都 骨 vs 骨

Simplified Chinese resulted in greater orthographic complexity

漢vs汉 頭vs头 樂vs乐 車vs车

- Minor glyph differences do not prevent mutual intelligibility
 写真 vs 写真 京都 vs 京都 骨 vs 骨
- Simplified Chinese resulted in greater orthographic complexity
 漢 vs 汉 頭 vs 头 樂 vs 乐 車 vs 车
- Many minor glyph differences can be treated as regional preferences

- Minor glyph differences do not prevent mutual intelligibility
 写真 vs 写真 京都 vs 京都 骨 vs 骨
- Simplified Chinese resulted in greater orthographic complexity
 漢 vs 汉 頭 vs 头 樂 vs 乐 車 vs 车
- Many minor glyph differences can be treated as regional preferences
- Writing systems undergo periodical reform

- Minor glyph differences do not prevent mutual intelligibility
 写真 vs 写真 京都 vs 京都 骨 vs 骨
- Simplified Chinese resulted in greater orthographic complexity
 漢 vs 汉 頭 vs 头 樂 vs 乐 車 vs 车
- Many minor glyph differences can be treated as regional preferences
- Writing systems undergo periodical reform
- There is much more cross-cultural digital communication than ever before



- Minor glyph differences do not prevent mutual intelligibility
 写真 vs 写真 京都 vs 京都 骨 vs 骨
- Simplified Chinese resulted in greater orthographic complexity
 漢 vs 汉 頭 vs 头 樂 vs 乐 車 vs 车
- Many minor glyph differences can be treated as regional preferences
- Writing systems undergo periodical reform
- There is much more cross-cultural digital communication than ever before
- Mutual intelligibility takes precedence over regional glyph preferences

- Minor glyph differences do not prevent mutual intelligibility
 写真 vs 写真 京都 vs 京都 骨 vs 骨
- Simplified Chinese resulted in greater orthographic complexity
 漢 vs 汉 頭 vs 头 樂 vs 乐 車 vs 车
- Many minor glyph differences can be treated as regional preferences
- Writing systems undergo periodical reform
- There is much more cross-cultural digital communication than ever before
- Mutual intelligibility takes precedence over regional glyph preferences
 - Some communication mediums are deemed more forgiving



- Minor glyph differences do not prevent mutual intelligibility
 写真 vs 写真 京都 vs 京都 骨 vs 骨
- Simplified Chinese resulted in greater orthographic complexity
 漢 vs 汉 頭 vs 头 樂 vs 乐 車 vs 车
- Many minor glyph differences can be treated as regional preferences
- Writing systems undergo periodical reform
- There is much more cross-cultural digital communication than ever before
- Mutual intelligibility takes precedence over regional glyph preferences
 - Some communication mediums are deemed more forgiving: *mobile*



- Minor glyph differences do not prevent mutual intelligibility
 写真 vs 写真 京都 vs 京都 骨 vs 骨
- Simplified Chinese resulted in greater orthographic complexity
 漢 vs 汉 頭 vs 头 樂 vs 乐 車 vs 车
- Many minor glyph differences can be treated as regional preferences
- Writing systems undergo periodical reform
- There is much more cross-cultural digital communication than ever before
- Mutual intelligibility takes precedence over regional glyph preferences
 - Some communication mediums are deemed more forgiving: *mobile*
- Some glyph differences are in (relatively) free variation within a single region

- Minor glyph differences do not prevent mutual intelligibility
 写真 vs 写真 京都 vs 京都 骨 vs 骨
- Simplified Chinese resulted in greater orthographic complexity
 漢 vs 汉 頭 vs 头 樂 vs 乐 車 vs 车
- Many minor glyph differences can be treated as regional preferences
- Writing systems undergo periodical reform
- There is much more cross-cultural digital communication than ever before
- Mutual intelligibility takes precedence over regional glyph preferences
 - Some communication mediums are deemed more forgiving: *mobile*
- Some glyph differences are in (relatively) free variation within a single region
- Han Unification is complex

Significant Events in Japan

Jōyō Kanji (1981)

Significant Events in Japan

- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added

Significant Events in Japan

- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji



- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)

- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji

- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (2010)



- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (2010)
 - 2,136 kanji in common use—196 kanji added; five kanji removed



- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (2010)
 - 2,136 kanji in common use—196 kanji added; five kanji removed
 - Incorporates many kanji from the NLC kanji set



- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (2010)
 - 2,136 kanji in common use—196 kanji added; five kanji removed
 - Incorporates many kanji from the NLC kanji set
 - Incorporates a mixture of simplified (pre-2010 revision) and traditional components



- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (2010)
 - 2,136 kanji in common use—196 kanji added; five kanji removed
 - Incorporates many kanji from the NLC kanji set
 - Incorporates a mixture of simplified (pre-2010 revision) and traditional components



- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (2010)
 - 2,136 kanji in common use—196 kanji added; five kanji removed
 - Incorporates many kanji from the NLC kanji set
 - Incorporates a mixture of simplified (pre-2010 revision) and traditional components

飯vs餅 慎vs塡

- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (2010)
 - 2,136 kanji in common use—196 kanji added; five kanji removed
 - Incorporates many kanji from the NLC kanji set
 - Incorporates a mixture of simplified (pre-2010 revision) and traditional components
 - 飯vs餅 慎vs塡 緑vs剝

- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (2010)
 - 2,136 kanji in common use—196 kanji added; five kanji removed
 - Incorporates many kanji from the NLC kanji set
 - Incorporates a mixture of simplified (pre-2010 revision) and traditional components

飯vs餅 慎vs塡 緑vs剝 込vs遡

- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (2010)
 - 2,136 kanji in common use—196 kanji added; five kanji removed
 - Incorporates many kanji from the NLC kanji set
 - Incorporates a mixture of simplified (pre-2010 revision) and traditional components

飯vs餅 慎vs塡 緑vs剝 込vs遡 著vs箸

- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (2010)
 - 2,136 kanji in common use—196 kanji added; five kanji removed
 - Incorporates many kanji from the NLC kanji set
 - Incorporates a mixture of simplified (pre-2010 revision) and traditional components

飯vs餅 慎vs塡 緑vs剝 込vs遡 著vs箸 謹vs僅

- Jōyō Kanji (1981)
 - 1,945 kanji in common use—95 kanji added
 - Basic principle: simplified forms are used only within the scope of Jōyō Kanji
- NLC (*National Language Council*) kanji set (2000)
 - Established the "preferred forms" for 1,022 kanji, above and beyond Jōyō Kanji
 - The preferred form is Japan's interpretation of the Kāngxī Zìdiǎn (康熙字典) form
- Jōyō Kanji (2010)
 - 2,136 kanji in common use—196 kanji added; five kanji removed
 - Incorporates many kanji from the NLC kanji set
 - Incorporates a mixture of simplified (pre-2010 revision) and traditional components

飯vs餅 慎vs塡 緑vs剝 込vs遡 著vs箸 謹vs僅 狭vs頰

Unification of regional glyph preferences into a single code point

- Unification of regional glyph preferences into a single code point
 - Han Unification

- Unification of regional glyph preferences into a single code point
 - Han Unification
 - Sets the stage for a "single glyph per code point" model

- Unification of regional glyph preferences into a single code point
 - Han Unification
 - Sets the stage for a "single glyph per code point" model
- Remember the exceptions!

- Unification of regional glyph preferences into a single code point
 - Han Unification
 - Sets the stage for a "single glyph per code point" model
- Remember the exceptions!
 - Source Separation Rule

- Unification of regional glyph preferences into a single code point
 - Han Unification
 - Sets the stage for a "single glyph per code point" model
- Remember the exceptions!
 - Source Separation Rule
 - Simplified forms

- Unification of regional glyph preferences into a single code point
 - Han Unification
 - Sets the stage for a "single glyph per code point" model
- Remember the exceptions!
 - Source Separation Rule
 - Simplified forms
- Some Simplified Chinese forms have been unified



- Unification of regional glyph preferences into a single code point
 - Han Unification
 - Sets the stage for a "single glyph per code point" model
- Remember the exceptions!
 - Source Separation Rule
 - Simplified forms
- Some Simplified Chinese forms have been unified
 - 骨 vs 骨



- Unification of regional glyph preferences into a single code point
 - Han Unification
 - Sets the stage for a "single glyph per code point" model
- Remember the exceptions!
 - Source Separation Rule
 - Simplified forms
- Some Simplified Chinese forms have been unified

骨 vs 骨 — Unified as U+9AA8

- Unification of regional glyph preferences into a single code point
 - Han Unification
 - Sets the stage for a "single glyph per code point" model
- Remember the exceptions!
 - Source Separation Rule
 - Simplified forms
- Some Simplified Chinese forms have been unified
 - 骨 vs 骨 Unified as U+9AA8

角 vs 角



- Unification of regional glyph preferences into a single code point
 - Han Unification
 - Sets the stage for a "single glyph per code point" model
- Remember the exceptions!
 - Source Separation Rule
 - Simplified forms
- Some Simplified Chinese forms have been unified
 - 骨 vs 骨 Unified as U+9AA8
 - 角 vs 角 Unified as U+89D2



• Future generations see regional glyph preferences as a communication barrier

- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model

- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model
 - Established a "single glyph per code point" model

- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model
 - Established a "single glyph per code point" model
 - Some amount of extrapolation was necessarily applied

- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model
 - Established a "single glyph per code point" model
 - Some amount of extrapolation was necessarily applied

- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model
 - Established a "single glyph per code point" model
 - Some amount of extrapolation was necessarily applied

鵤→鵤 (U+9D64)

• The original (2000) version did this for the URO and Extension A in their entirety

- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model
 - Established a "single glyph per code point" model
 - Some amount of extrapolation was necessarily applied

- The original (2000) version did this for the URO and Extension A in their entirety
- The 2005 revision expanded this to include Extension B in its entirety



- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model
 - Established a "single glyph per code point" model
 - Some amount of extrapolation was necessarily applied

- The original (2000) version did this for the URO and Extension A in their entirety
- The 2005 revision expanded this to include Extension B in its entirety
- The only national standard that includes entire CJK Unified Ideographs blocks



- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model
 - Established a "single glyph per code point" model
 - Some amount of extrapolation was necessarily applied

- The original (2000) version did this for the URO and Extension A in their entirety
- The 2005 revision expanded this to include Extension B in its entirety
- The only national standard that includes entire CJK Unified Ideographs blocks
- The big question:



- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model
 - Established a "single glyph per code point" model
 - Some amount of extrapolation was necessarily applied

- The original (2000) version did this for the URO and Extension A in their entirety
- The 2005 revision expanded this to include Extension B in its entirety
- The only national standard that includes entire CJK Unified Ideographs blocks
- The big question: Which regional glyph becomes the preferred glyph?



- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model
 - Established a "single glyph per code point" model
 - Some amount of extrapolation was necessarily applied

- The original (2000) version did this for the URO and Extension A in their entirety
- The 2005 revision expanded this to include Extension B in its entirety
- The only national standard that includes entire CJK Unified Ideographs blocks
- The big question: Which regional glyph becomes the preferred glyph?
- One could argue that GB 18030 already achieved Genuine Han Unification

- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model
 - Established a "single glyph per code point" model
 - Some amount of extrapolation was necessarily applied

- The original (2000) version did this for the URO and Extension A in their entirety
- The 2005 revision expanded this to include Extension B in its entirety
- The only national standard that includes entire CJK Unified Ideographs blocks
- The big question: Which regional glyph becomes the preferred glyph?
- One could argue that GB 18030 already achieved Genuine Han Unification
 - There was obvious—and necessary—bias in the choice of glyphs

- Future generations see regional glyph preferences as a communication barrier
- GB 18030 offers a glimpse into a possible model
 - Established a "single glyph per code point" model
 - Some amount of extrapolation was necessarily applied

- The original (2000) version did this for the URO and Extension A in their entirety
- The 2005 revision expanded this to include Extension B in its entirety
- The only national standard that includes entire CJK Unified Ideographs blocks
- The big question: Which regional glyph becomes the preferred glyph?
- One could argue that GB 18030 already achieved Genuine Han Unification
 - There was obvious—and necessary—bias in the choice of glyphs
- Motivation and international cooperation are key success factors

Is Genuine Han Unification Possible?

• I firmly believe so, but it is not possible today


- I firmly believe so, but it is not possible today
 - There exists far too much bias and preference in today's generations and regions

- I firmly believe so, but it is not possible today
 - There exists far too much bias and preference in today's generations and regions
- My prediction:

- I firmly believe so, but it is not possible today
 - There exists far too much bias and preference in today's generations and regions
- My prediction: Genuine Han Unification is possible within 25 years

- I firmly believe so, but it is not possible today
 - There exists far too much bias and preference in today's generations and regions
- My prediction: Genuine Han Unification is possible within 25 years
- There is sufficient evidence to back up this prediction

- I firmly believe so, but it is not possible today
 - There exists far too much bias and preference in today's generations and regions
- My prediction: Genuine Han Unification is possible within 25 years
- There is sufficient evidence to back up this prediction
 - Unicode serves as the foundation for modern digital text

- I firmly believe so, but it is not possible today
 - There exists far too much bias and preference in today's generations and regions
- My prediction: Genuine Han Unification is possible within 25 years
- There is sufficient evidence to back up this prediction
 - Unicode serves as the foundation for modern digital text
 - Periodic changes to the preferred form of region-specific glyphs

- I firmly believe so, but it is not possible today
 - There exists far too much bias and preference in today's generations and regions
- My prediction: Genuine Han Unification is possible within 25 years
- There is sufficient evidence to back up this prediction
 - Unicode serves as the foundation for modern digital text
 - Periodic changes to the preferred form of region-specific glyphs
 - Ever-increasing cross-cultural communication



- I firmly believe so, but it is not possible today
 - There exists far too much bias and preference in today's generations and regions
- My prediction: Genuine Han Unification is possible within 25 years
- There is sufficient evidence to back up this prediction
 - Unicode serves as the foundation for modern digital text
 - Periodic changes to the preferred form of region-specific glyphs
 - Ever-increasing cross-cultural communication
 - Users tend to be more forgiving for region-specific glyphs in mobile environments



- I firmly believe so, but it is not possible today
 - There exists far too much bias and preference in today's generations and regions
- My prediction: Genuine Han Unification is possible within 25 years
- There is sufficient evidence to back up this prediction
 - Unicode serves as the foundation for modern digital text
 - Periodic changes to the preferred form of region-specific glyphs
 - Ever-increasing cross-cultural communication
 - Users tend to be more forgiving for region-specific glyphs in mobile environments
 - This may lead people to eventually becoming "numb" to such differences



"Genuine Han Unification is not outside the realm of extreme possibilities." — Fox William Mulder, FBI Special Agent

