

今さら聞けない文字コードのはなし

2019年11月19日

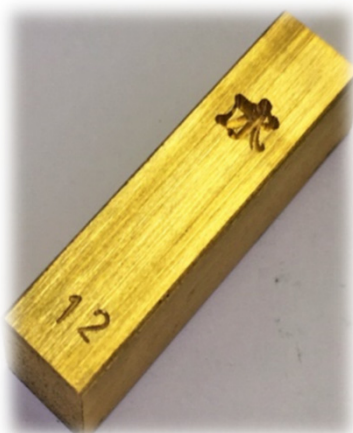
株式会社 **イワタ**

水野 昭

本資料は2018年11月29日に東京ビッグサイトで開催された
JANPS2018（新聞製作技術展・日本新聞協会主催）の会場内
セミナーで発表した資料をもとにしています。

本資料の無断転載・無断使用を禁止します。

イワタの歴史



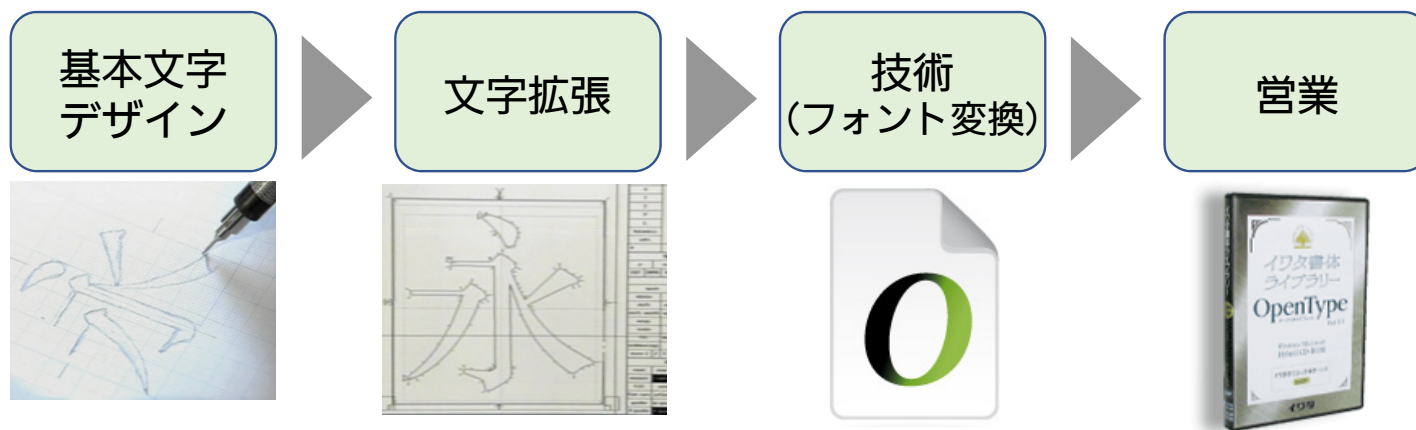
- 1920年（大正9） 岩田百蔵が創業。鉛活字の母型を製造。
- 1939年（昭和14） 大阪龍文堂の森川健市（のちの大阪支店長）が新聞用扁平活字を開発
- 1950年（昭和25） ベントン彫刻機導入をきっかけに日本最大の母型製造工場に成長。
- 1988年（昭和63） デジタル部門設立
- 2006年（平成18） イワタUDフォントをパナソニックと共同開発
- 2007年（平成19） イワタUDフォントが日本新聞協会技術開発奨励賞を受賞
- 2009年（平成21） イワタUD新聞明朝を信濃毎日新聞社と共同開発
- 2015年（平成27） イワタミンゴNPを発売



イワタの特徴

- 電機機器、印刷業からの**受託開発（特注フォント）**が中心。
全体売り上げの半分以上を占める
- 書体開発に必要なスタッフ（文字デザイン、文字拡張、技術、営業）
をすべて持つ。日本のフォントベンダーでは珍しい。

< フォント制作の流れ >



イワタの代表書体

■イワタ明朝体オールド

活字時代から使われる明朝体オールド

■イワタ新聞明朝体]

新聞明朝体はフトコロが広いのが特徴です

■イワタ教科書体]

教科書体は主に小学校で使われます。

■イワタUDフォント]

横組み専用の UD フォント

新聞社との長いおつきあい

活字時代には100社超の新聞社様と取り引きしていました（1950～1992年の期間）

DS社 日本農業新聞社 東京ニュース写真製版 第一新聞印刷 日刊スポーツ社 日本証券新聞社 日本経済新聞社 日本機関紙印刷 東京タイムズ印刷社 東京新聞社 代々木あかつき印刷 株式新聞社 大東新聞印刷 燃料油脂新聞 読売新聞社 読売報知印刷所 中部読売新聞社 読売新聞社西部本社 日本新聞印刷所 新聞之新聞社 新日本印刷 日刊工業新聞社 新日本印刷 朝日新聞社 東京商品日報社 産報印刷 住宅新報 時事新報社 新橋新聞印刷所 産経新聞社 生産経済新聞社 自由新聞 新橋新聞印刷 油業報知新聞 東京新聞印刷工業会 滝沢新聞印刷 日経印刷 城東新聞印刷 毎日新聞社 織研新聞社 紙業日々新聞 自動車週報 鹿児島新報社 夕刊フクニチ 米沢新聞社 福島民友新聞社 茨城新聞社 石巻新聞社 岩手日々新聞社 岩手日報社 浜通新聞社 西日本新聞社 西日本新聞印刷 北海道新聞社 北国新聞社 東奥日報社 栃木新聞 中日新聞北陸本社 千葉日報社 合同新聞社 神奈川新聞社 河北新報社 陸奥新報社 室蘭民法 熊本日々新聞社 山梨日々新聞 新いばらきタイムス社 山梨時事 山形新聞社 福井新聞社 デーリー東北新聞社 佐賀新聞社 埼玉新聞社 埼玉新聞社 大衆日報社 神静民報社 秋田魁新報社 山陽新聞社 東方新聞印刷 長崎時事 北日本新聞社 南日本新聞社 四国新聞社 荘内日報社 信濃毎日新聞社 静岡新聞社 上毛新聞社 日向日々新聞社 常陽新聞社 山陰日々新聞社 大阪商要新聞 アサヒスタチオ 青森毎日新聞社 日本食糧新聞社 機関紙印刷 神戸新聞社 繊維新聞 防長新聞社 みなと新聞 西日本木材新聞社 名古屋タイムス印刷 小池製作所 神戸新聞社 房総毎日

現在のイワタ

現在は、全国約50の新聞社様と取り引き中。

「新聞をより読みやすく・より伝わりやすく」

なるよう書体開発に意欲的に取り組んでいます。

[代表書体]

イワタUD新聞明朝

イワタUD新聞ゴシック

イワタミンゴNP

イワタブースで実際の紙面をご覧ください。

本セミナーを行うきっかけ

■ 文字と書体について知る人が少なくなった

- － 会社の合理化、世代交代
- － 地方紙ではシステム部門が文字兼任
- － 大手紙のフォント部門は縮小傾向

■ 文字と書体について勉強したいという声が多い

- － 勉強する機会がない、書籍も少ない
- － コンピュータの日本語環境は常に進化している
- － 新聞制作に文字（特に文字コード）の知識は必要不可欠

このような話は新聞業界に限りません

本セミナーの内容

■ 文字コード、特にUnicodeについてお話しします

- ー Unicodeが最大にして最新の文字コード体系
- ー 現在のコンピュータ（OS、アプリ）はUnicodeで動いている
- ー U-PRESSもUnicodeをベース

■ Unicodeの最新情報

■ U-PRESSの特徴と課題

いろいろな文字集合

1. ASCII

初期のパソコンで使用された最も普及しているコード体系
7bit使用し128文字使用可能。制御文字などを除き94文字を割り当て

2. ISO 646

ISO（国際標準化機構）がASCIIを国際化したもの
\$ [/]などASCIIの一部(12文字)を各国で入替可能にしたもの

3. JISX0201

ISO646をJIS（日本工業規格）化したもので半角カタカナを追加
ASCIIのx5c（バックスラッシュ）を円マーク¥にしている

4. JISX0208

日本が作った世界最初が多バイト文字コード
90年版を最後に字数と字形に変更がないことから**90JIS**などと呼ばれる

文字を区点コードに割り当てし全6879文字収容
漢字は使用頻度に応じて**第1水準漢字**、**第2水準漢字**に分かれる
（第1水準は五十音順、第2水準は部首画数順に収容）

いろいろな文字集合

5. JISX0213

2000年に策定。

JISX0208を拡張するための第3水準文字、第4水準文字から構成される。

JISX0208と一緒に使うことが前提

JISX0208の区点コードと重複しないように文字を割り当てている
従来の「区点」では文字を収容できず、新たに「面」の概念を追加。

2004年に改訂され**04JIS**とも呼ばれる。

JISX0208およびJISX0213の**168文字の字形変更**と10文字の追加。

現在の日本語を公に定義する最新・最大の文字セット

JIS区点概念図

2000年以前まで日本語は
JISX0208
JISX0212（補助漢字）
という2つの独立した文字集合があった

JISに関わった学者、専門家は
この2つの区点表（最大8836文字×2）に
必要文字を埋めていった。

収容文字数

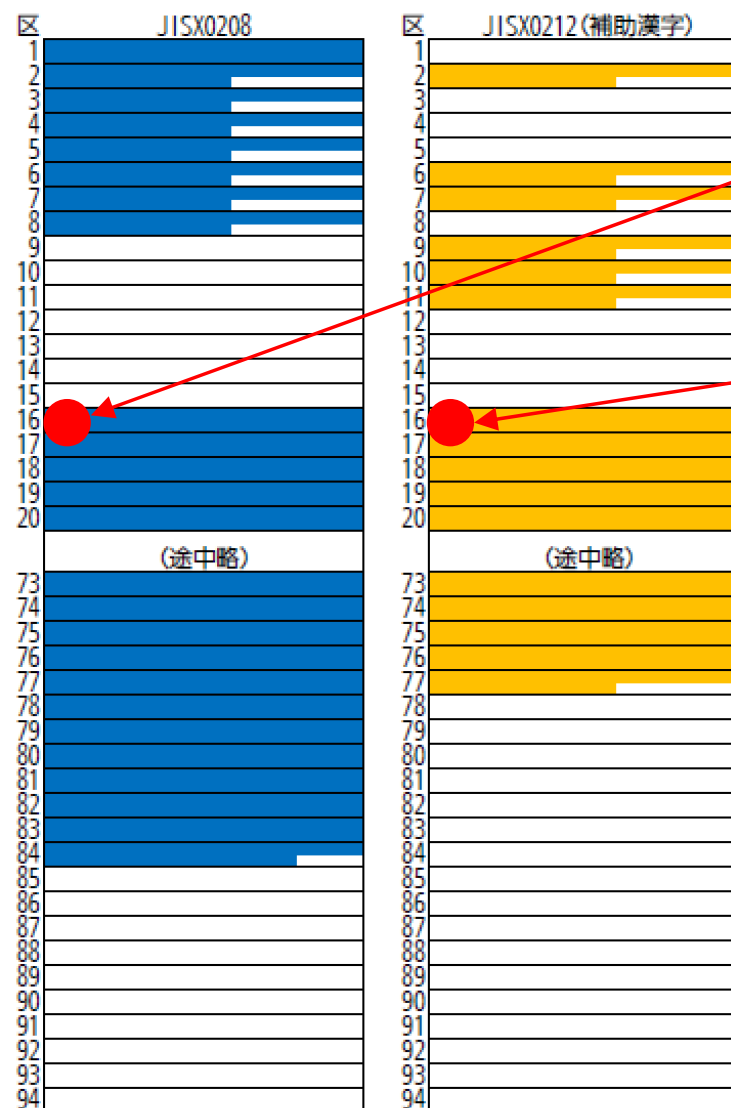
JISX0208：6879文字（漢6355＋非524）

JISX0212：6067文字（漢5801＋非266）

JISX0212はJISX0208にない情報処理系文字
を収録している（アルファベットや記号あり）

■ JISX0208

■ JISX0212（補助漢字）



亜

JISX0208
16区01点

𠩺

JISX0212
16区01点

JISX0213:2000

2000年にJISX0213を策定

JISX0208とJISX0212と重複しないよう
2枚の区点表のすきまに文字を収容

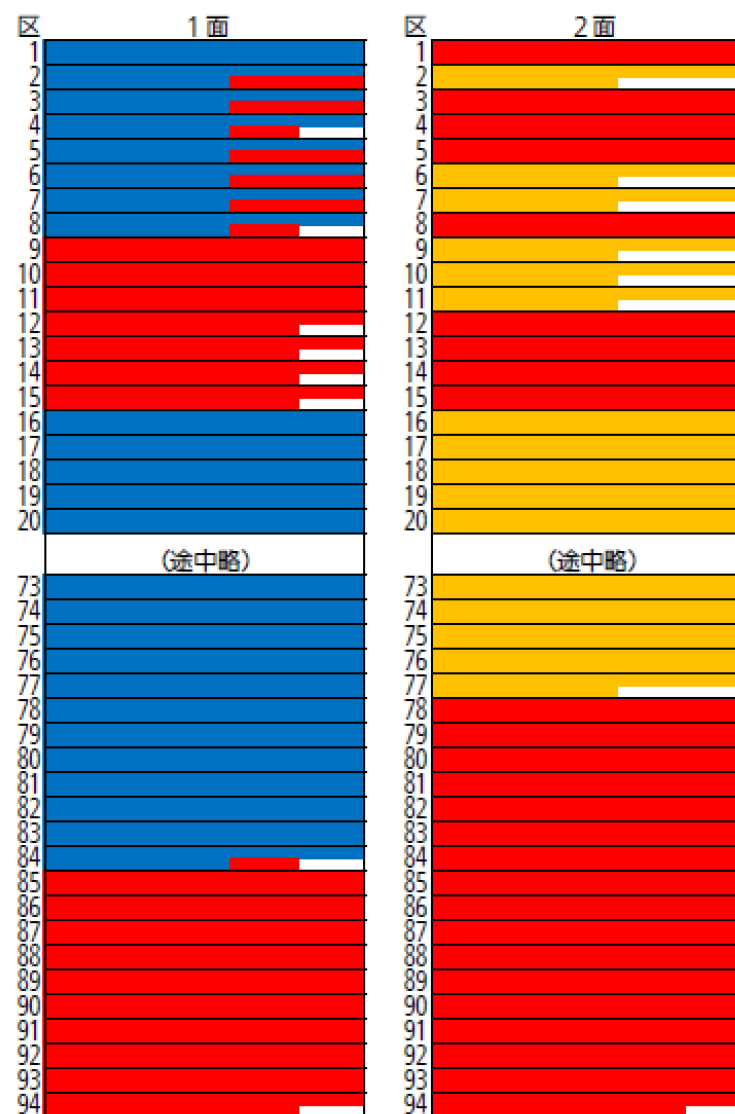
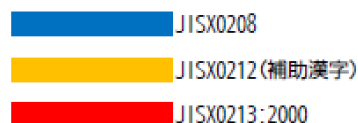
2つの区点表を区別するため「面」を導入

第1面に追加した文字を第3水準
第2面に追加した文字を第4水準

収容文字数

JISX0213 : 4344文字（漢3685＋非659）

JISX0213は主に印刷文字を収録している



JISX0213:2004

2004年にJISX0213が改訂

第1面（第3水準）に10文字追加され
さらに168文字の例示字形が変更された

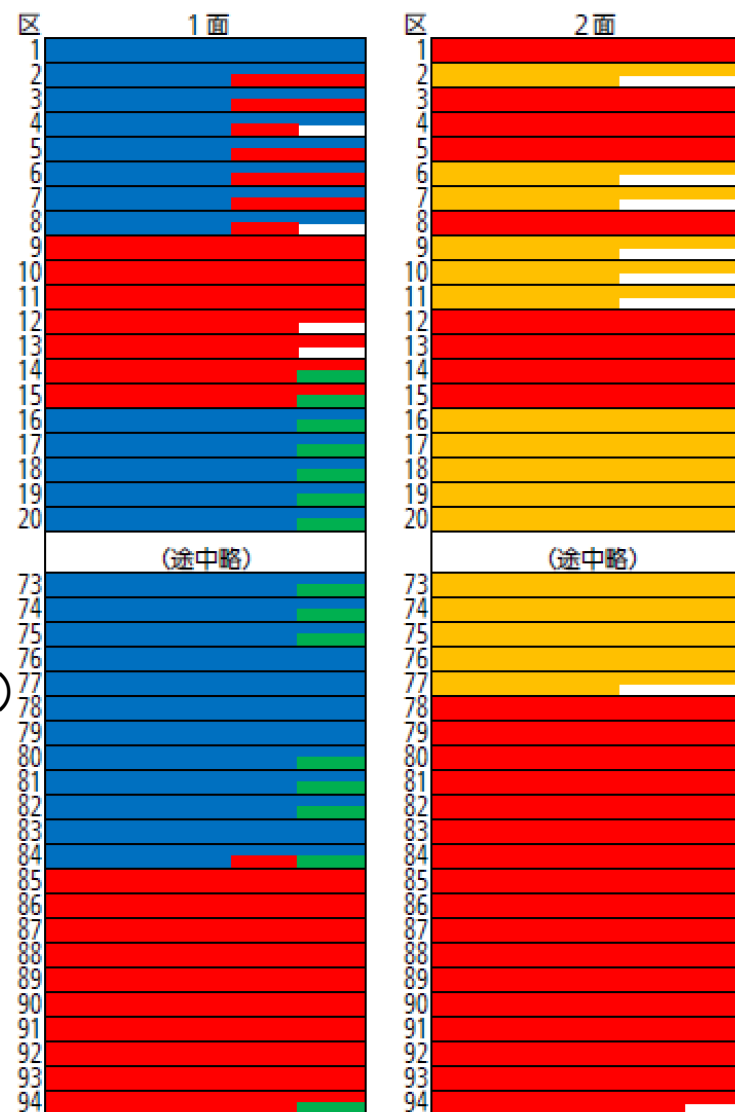
この規格を04JISと呼ぶことがある
またこの変更された文字を**04字形**と呼ぶ

収容文字数

JISX0208 : 6879文字（漢6355＋非524）

JISX0213 : 4354文字（漢3695＋非659）

計11233文字（漢10050＋非1183）



話題になった04字形

2004年に例示字形が変更された文字の一部（左が90字形,右が04字形）

逢→逢	祇→祇
茨→茨	薩→薩
鰯→鰯	鯖→鯖
樽→樽	杖→杖
葛→葛	賭→賭

U-PRESS移行後も個別に字形を
戻す社もあります

例) 祇園 薩摩 逢う

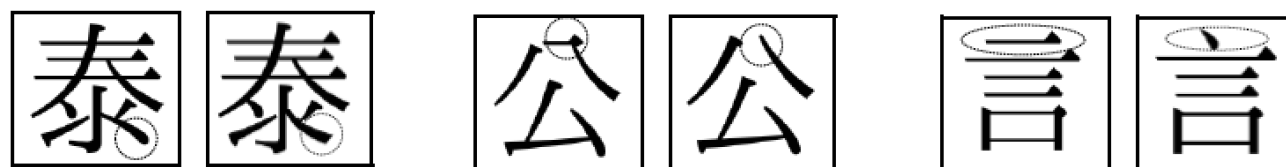
WindowsVISTAでバンドルフォントの字形が04字形になり話題になった

U-PRESSは右の04字形を採用

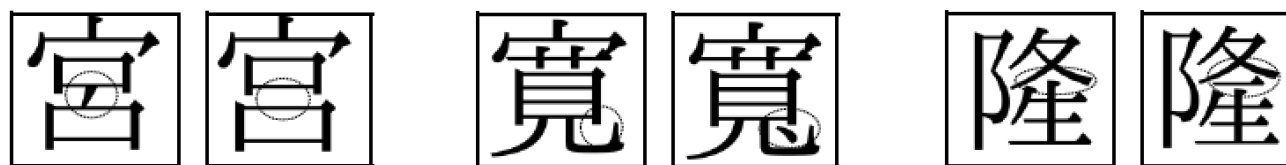
JISとしてはどちらの文字も包摂の範囲内でありどちらを用いても良い
書体によっても字形は異なる場合がある（ゴシックや筆書系）

字形に対するの許容差は人それぞれ

筆法による字形差は許容できる人が多い



画数に差がある文字は許容できない人が多い



引用元)

コンピュータマインド株式会社 (川崎市宮前区)

2017年「字形に対する意識について～漢字の形に関するアンケート結果から～」

Unicode（ユニコード）とは

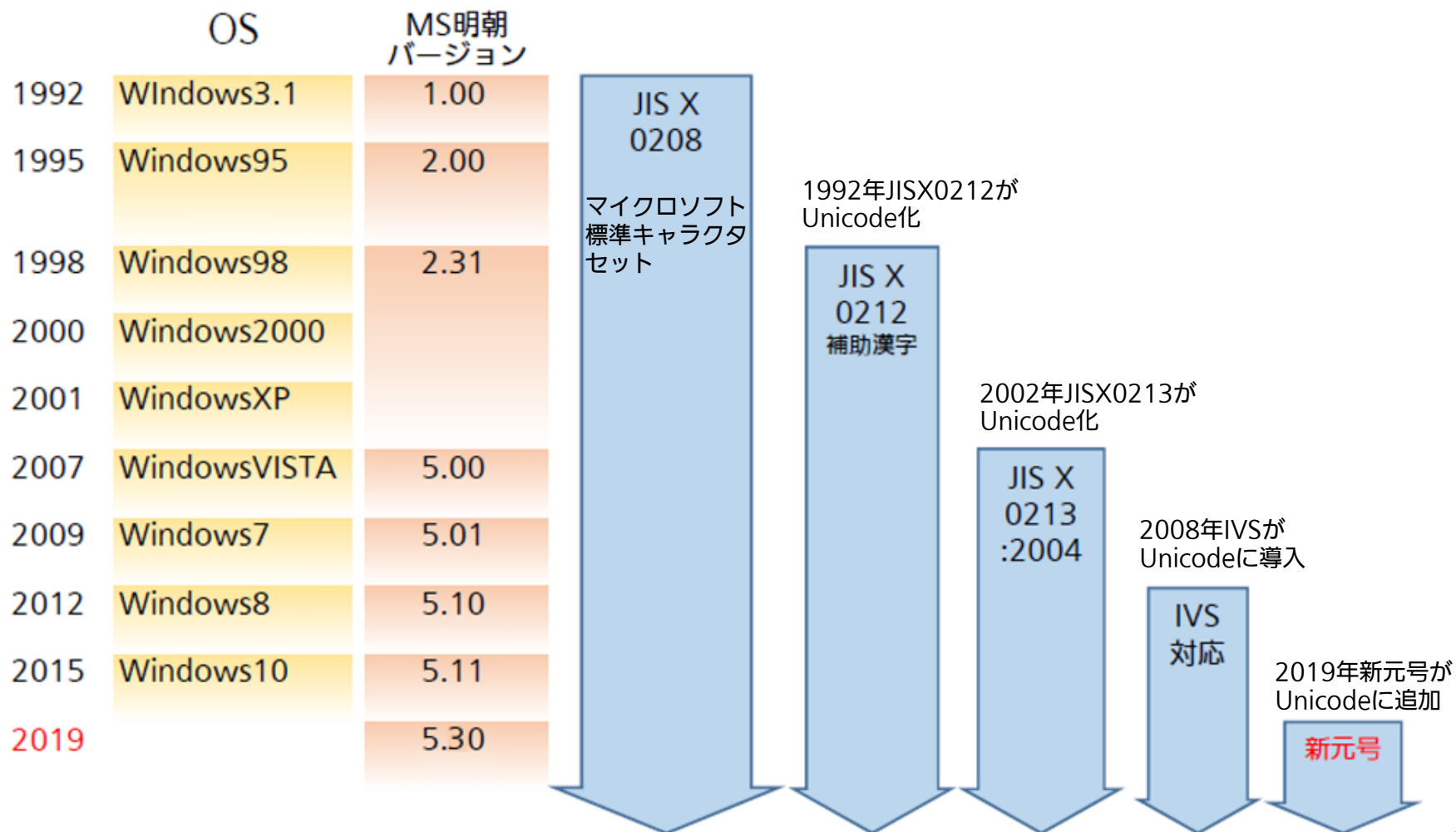
Unicodeとは世界の全言語収容を目標とした文字セット。
現在のコンピュータ（OS、アプリ）は文字のやりとりにUnicodeを使用

- Unicode
1990年にApple, Xerox, Microsoft, IBMなどの企業が作った文字集合
Unicodeコンソーシアムという組織が管理
- ISO/IEC 10646
Unicodeを国際規格化したもの
(Unicodeを追認している)
- JISX0221
ISO/IEC 10646をJIS化したもの

Unicodeの変遷

制定年	バージョン	文字数	概要
1991年	Unicode 1.0.0	7,161	16ビット, JISX0201
1992年	Unicode 1.0.1	28,359	JISX0208, JISX0212
1996年	Unicode 2.0.0	38,950	サロゲートペア
1998年	Unicode 2.1.0	38,952	ユーロ記号
1999年	Unicode 3.0.0	49,259	CJK統合漢字拡張A, AJ1-4
2001年	Unicode 3.1.0	94,205	CJK統合漢字拡張B (第2面を使用)
2002年	Unicode 3.2.0	95,221	JISX0213, AJ1-5
2003年	Unicode 4.0.0	96,447	
2006年	Unicode 5.0.0	99,089	
2008年	Unicode 5.1.0	100,713	異体字セレクタ(IVS), AJ1-6 (U-PRESS)
2009年	Unicode 5.2.0	107,361	ARIB外字
2010年	Unicode 6.0.0	109,449	携帯電話の絵文字
2014年	Unicode 7.0.0	113,021	
2015年	Unicode 8.0.0	120,737	
2016年	Unicode 9.0.0	128,172	カラー文字, 4KTV放送用シンボル
2017年	Unicode 10.0.0	136,690	変体仮名285, 文字情報基盤完了 (漢字拡張F)
2018年	Unicode 11.0.0	137,374	
2019年3月	Unicode 12.0.0	137,928	
2019年5月	Unicode 12.1.0	137,929	令和合字 (U+32FF)

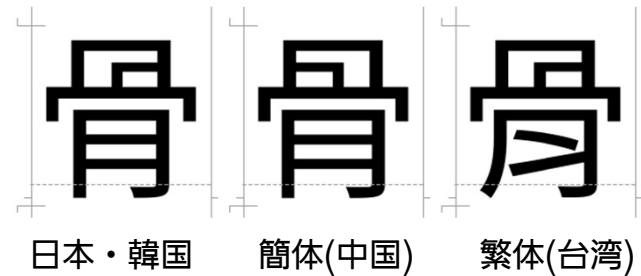
Windowsと対応文字



CJK統合漢字の採用

日本・中国・韓国の漢字の類似する文字を統合
(形状が似ていて由来が同じ文字は同一コードとするのが原則)

Unicode 9AA8の例



当初16ビットで全文字を扱うことを目指した
(16ビット：0000～FFFFの計65536文字)

のちに16ビットで収まらないことが明らかとなり、
32ビット領域に統合漢字の拡張部を設けることになる

Unicodeは16ビットから32ビットへ

16ビットの世界

6 1 1 B 愛

32ビットの世界

0 0 0 0 6 1 1 B 愛

611Bも0000611Bも指している文字コードは同じ

32ビットで表現できる16進数は

00000000～FFFFFFFF

これで最大43億文字が表現できるが、
実際には1bitを別用途に使っているので21億文字まで

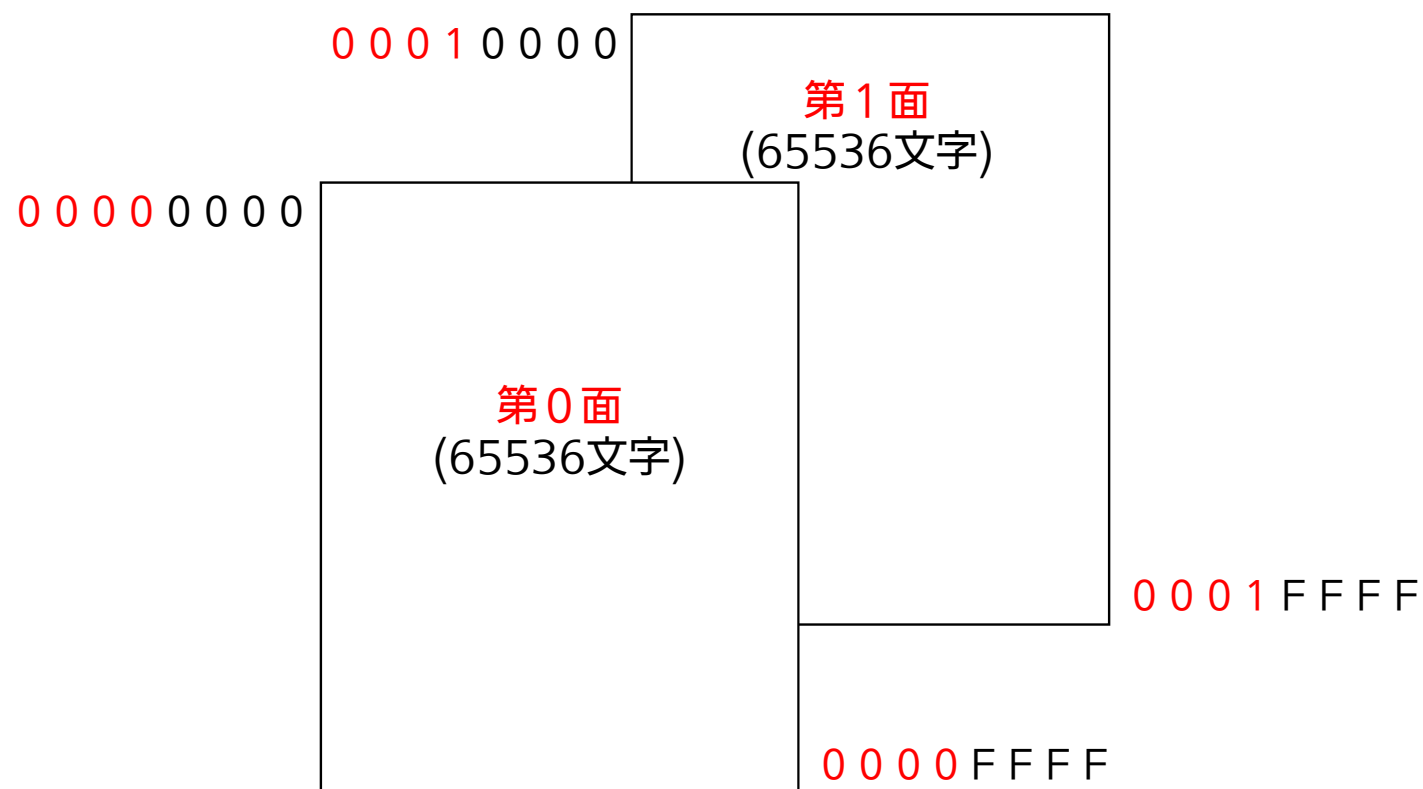
16ビット文字集合のイメージ

0 0 0 0 ~ F F F F を 1 枚の表 (=面) に見立てると以下になる
初期のUnicode (バージョン1) はこんな感じ



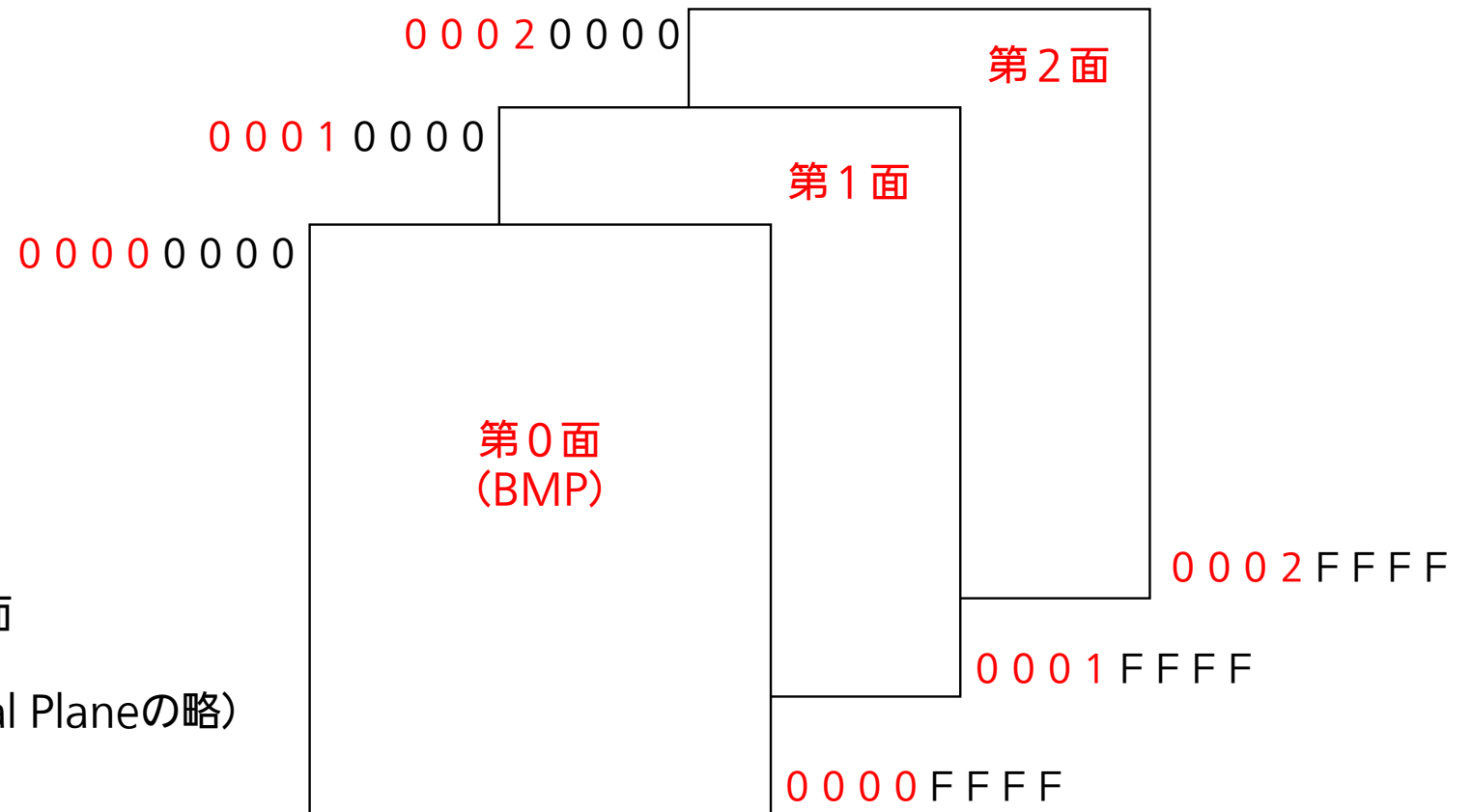
多面化文字集合のイメージ

16ビットで足りなくなったのでもう1面追加し32ビットで管理する



多面化文字集合のイメージ

現在のUnicodeで文字の割り当てがあるのは第2面まで



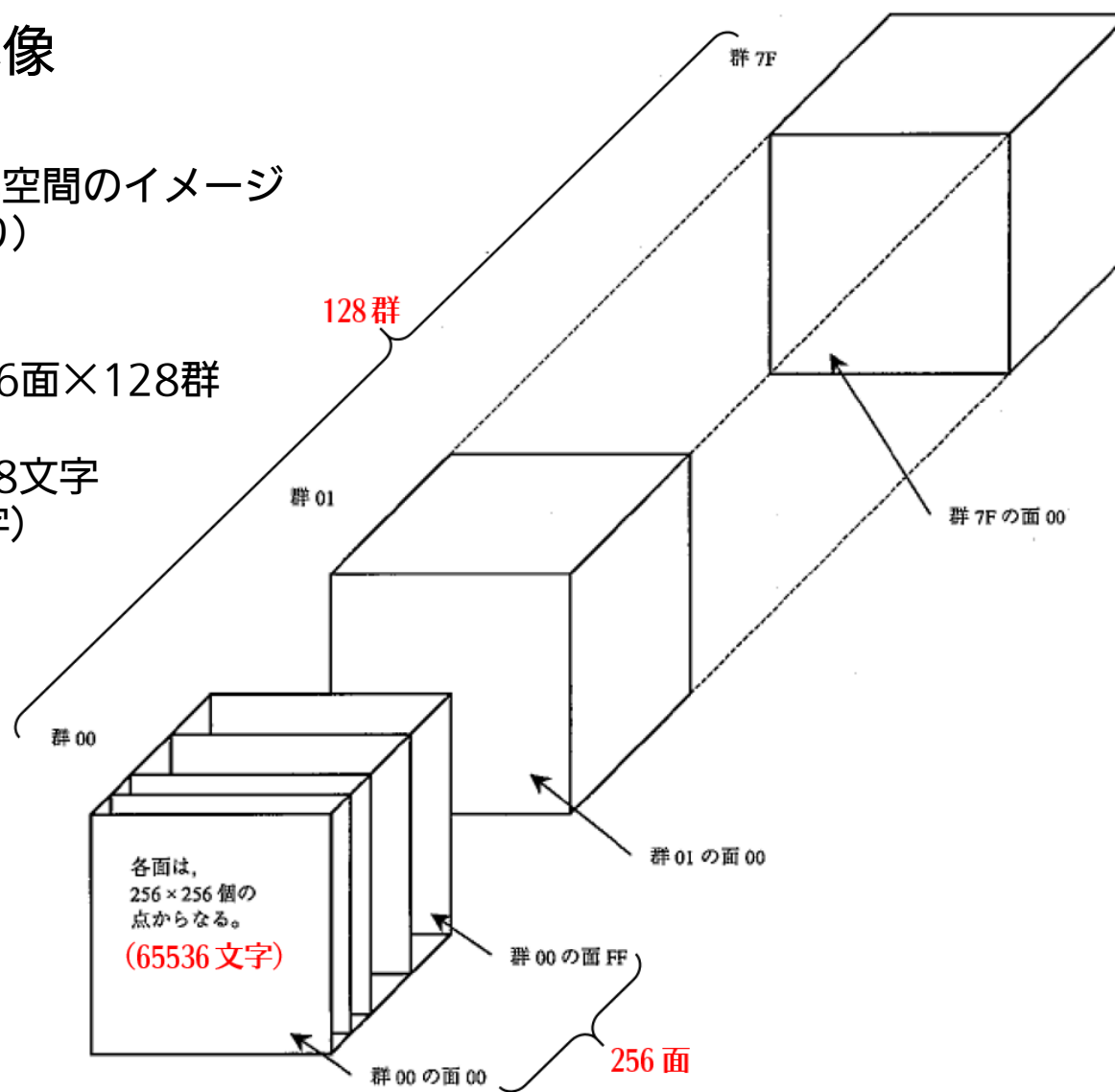
第0面を基本多言語面
またはBMPという
(Basic Multilingual Planeの略)

Unicodeの全体像

UCS-4の全符号化空間のイメージ
(ISO-10646より)

65,536文字×256面×128群

||
2,147,483,648文字
(約21億文字)



各面の用途

現在のUnicodeはでは32,768面（256面×128群）のうち
6面のみ割り当てが決まっている

0 0 0 0 0 0 0 0 ~ 0 0 0 0 F F F F	...	Unicode 0 面（文字収容）
0 0 0 1 0 0 0 0 ~ 0 0 0 1 F F F F	...	Unicode 1 面（文字収容）
0 0 0 2 0 0 0 0 ~ 0 0 0 2 F F F F	...	Unicode 2 面（文字収容）
0 0 0 E 0 0 0 0 ~ 0 0 0 E F F F F	...	Unicode 14 面（IVS用）
0 0 0 F 0 0 0 0 ~ 0 0 0 F F F F F	...	Unicode 15 面（私用領域）
0 0 1 0 0 0 0 0 ~ 0 0 1 0 F F F F	...	Unicode 16 面（私用領域）

登録された文字が入っているのは0, 1, 2 面
（一部サロゲートペア用のエリアあり）

14面は異体字セレクタ（IVS）用
15, 16面は私用領域（外字エリア）

上記以外のエリアは今のところ未定義

漢字の収容エリア

面	Unicode			項目	説明
第0面 (BMP)	3400	～	4DB5	CJK統合漢字 拡張A ②	Unicode3.0で追加の統合漢字(6582文字)
	4E00	～	9FEA	CJK統合漢字 ①	中国、日本、韓国の文字を統合した漢字(20976文字)
	E000	～	F8FF	私用領域	6400文字分
	F900	～	FAFF	CJK互換漢字	
第2面	20000	～	2A6D6	CJK統合漢字 拡張B ③	Unicode3.1で追加の統合漢字(42711文字)
	2A700	～	2B734	CJK統合漢字 拡張C ④	Unicode5.2で追加の統合漢字(4149文字)
	2B740	～	2B81D	CJK統合漢字 拡張D ⑤	Unicode6.0で追加の統合漢字(222文字)
	2B820	～	2CEA1	CJK統合漢字 拡張E ⑥	Unicode8.0で追加の統合漢字(5762文字)
	2CEB0	～	2EBE0	CJK統合漢字 拡張F ⑦	Unicode10.0で追加の統合漢字(7473文字)
	2F800	～	2FA1F	CJK互換漢字補助	
第14面	E0100	～	E01EF	VS(VariationSelectors)	異体字選択用
第15面	F0000	～	FFFFF	私用領域A	
第16面	100000	～	10FFFF	私用領域B	

サロゲートペア

16ビット文字コードを2つ使用して1文字を表現する方法

文字情報

文字(M) 丈

異体字(B)

音読み: ジョウ

訓読み:

部首読み: いち

文字画数: 4

常用/人名: 該当しない

漢字配当: 配当外

漢字水準: 第3水準

JIS: シフトJIS:

区点: 面区点: 1-14-02

Unicode: U+2000B(D840 DC0B)

UTF-8: F0 A0 80 8B

確定(N) コピー(C) 閉じる

これ

UTF-32だと2000B

UTF-16だとD840 DC0B

UTF-8だとF0 A0 80 8B

ATOK文字パレット

UCSとUTF

■ UCS

文字を入れる**器の大きさ**のこと

(Universal multipul-octet coded Character Setの略)

UCS-2 : 2バイトで表現できる文字集合

2バイト=16ビット=65,536文字

UCS-4 : 4バイトで表現できる文字集合

4バイト=32ビット=約43億文字

しかし32ビットのうち最上位ビットは0固定のため約21億文字

■ UTF

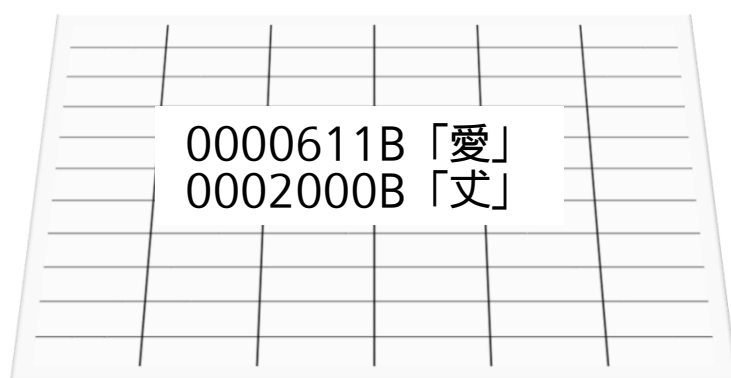
コンピュータがUCSから文字を取り出し**伝送するための手法**

(UCS Transformation Formatの略)

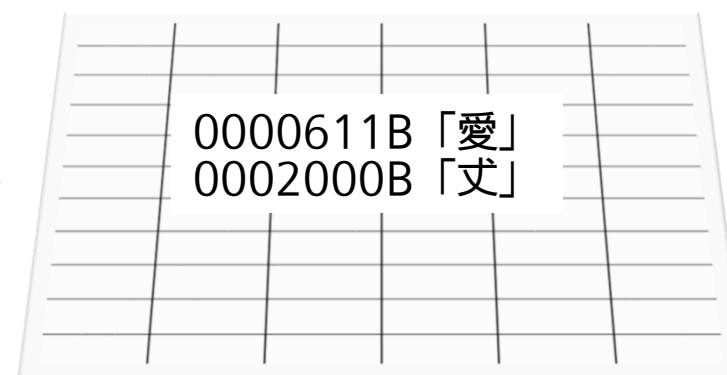
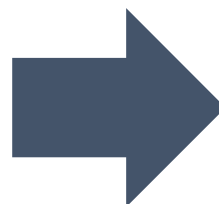
UTF-16 : 16ビットを1単位として伝送

UTF-32 : 32ビットを1単位として伝送

Unicode 2 面文字の伝送方法 [UTF-32の場合]



UCS-4の文字集合



「愛」も「丈」も問題なし

サロゲートペア

第0面（BMP）の計65,536文字のうち、
以下のエリアに文字を割り当てず**組み合わせ専用の領域**とする

[サロゲート領域]

D800～DBFF（1024コード）

DC00～DFFF（1024コード）

ここを使うと $1,024 \times 1,024 = 1,048,576$ 文字がさらに追加で表現可能となる

結局UTF-16で表現可能な文字数は
 $63,488 \text{文字} + 1,048,576 \text{文字} = 1,112,064 \text{文字}$

第0面から第16面の計17面分に相当

「丈」 0x0002000b はUTF-16では U+D840 U+DC0B

「𪛗」 0x00020B9F はUTF-16では U+D842 U+DF9F

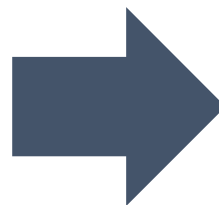
Unicode 2 面文字の伝送方法 [UTF-16の場合]

サロゲートペアを使う場合
(16ビットを2回用いると)

A diagram showing a 6x6 grid representing UCS-4. The first two rows contain the hexadecimal values 0000611B and 0002000B, followed by the characters '愛' and '丈' respectively.

				0000611B	「愛」
				0002000B	「丈」

UCS-4の文字集合



A diagram showing a 6x6 grid representing UTF-16. The first two rows contain the hexadecimal values 611B and D840+DC0B, followed by the characters '愛' and '丈' respectively.

				611B	「愛」
				D840+DC0B	「丈」

「愛」も「丈」も問題なく表示

変わってきた漢字追加のポリシー

追加される漢字

従来には無い字形で新規登録がふさわしいと認められた漢字には新たな文字コードが与えられる

字形が似ていて由来が同じ場合は同じ文字扱いというのが原則だった

異体字には？

従来、統合漢字として扱われていた異体字の中で、独立コードとして新規登録するほどの字形ではないものの、その必要性が認められた文字には親文字の枝番号が与えられ字形を区別することが可能となった

この枝番号を異体字セレクタ（IVS）といい親文字とセットで扱う
（IVS=Ideographic Variation Sequence）

2008年のUnicode5.1から導入

IVSとは

IVSとは

漢字をより細かいレベルで字形を表示・区別するためのしくみ

Adobeのソフトでしか異体字が使えない、メモ帳で異体字が使えないだったのが、環境が揃えば異体字が使えるようになる。

しくみ

親文字のUnicodeと枝番号（VS）のUnicodeを組み合わせて異体字を表現

VS : Variation Selector, UnicodeのE0100～E01EF（第14面）を使用

- VSはただの枝番号。親文字と組み合わせて始めて固有の字形となる
- 字形はUnicodeに登録する制度になっている
- 登録された一覧表をIVD（Ideographic Variation Database）という（通常、ある目的をもった文字集合＝コレクションごとに登録される）

IVSとは

- あいまいだった統合漢字字形がIVSで一意的に字形が定まる

845B + E0100 → 葛

葛 or 葛

845B + E0101 → 葛

葛 or 葛

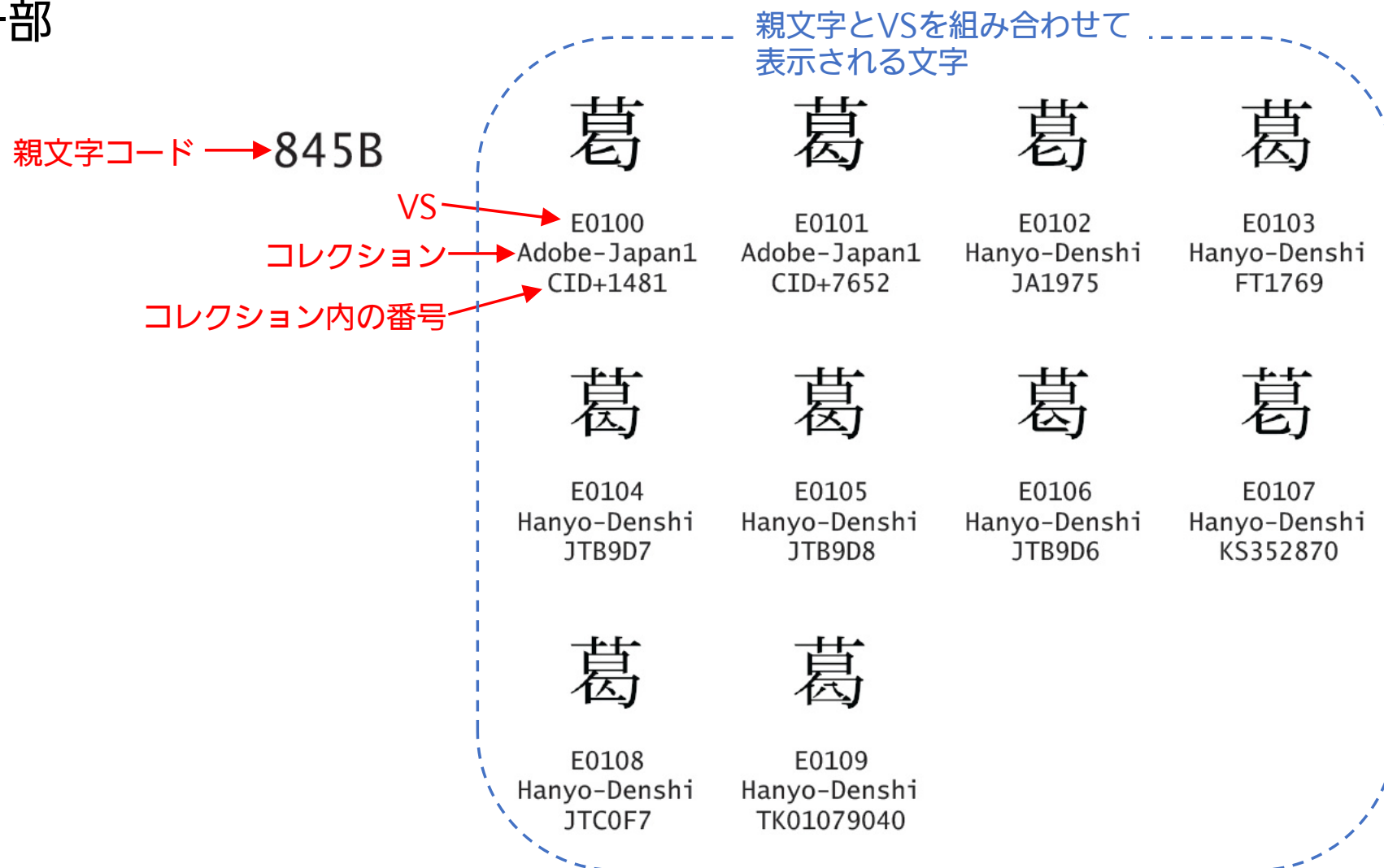
- 日本語に関する主なIVDコレクション

Adobe-Japan1 : アドビが登録。出版業界や印刷業界で多用。

Hanyo-Denshi : 経産省の委託事業「汎用電子情報交換環境整備プログラム」で
戸籍文字と住基ネット文字を整理したもの

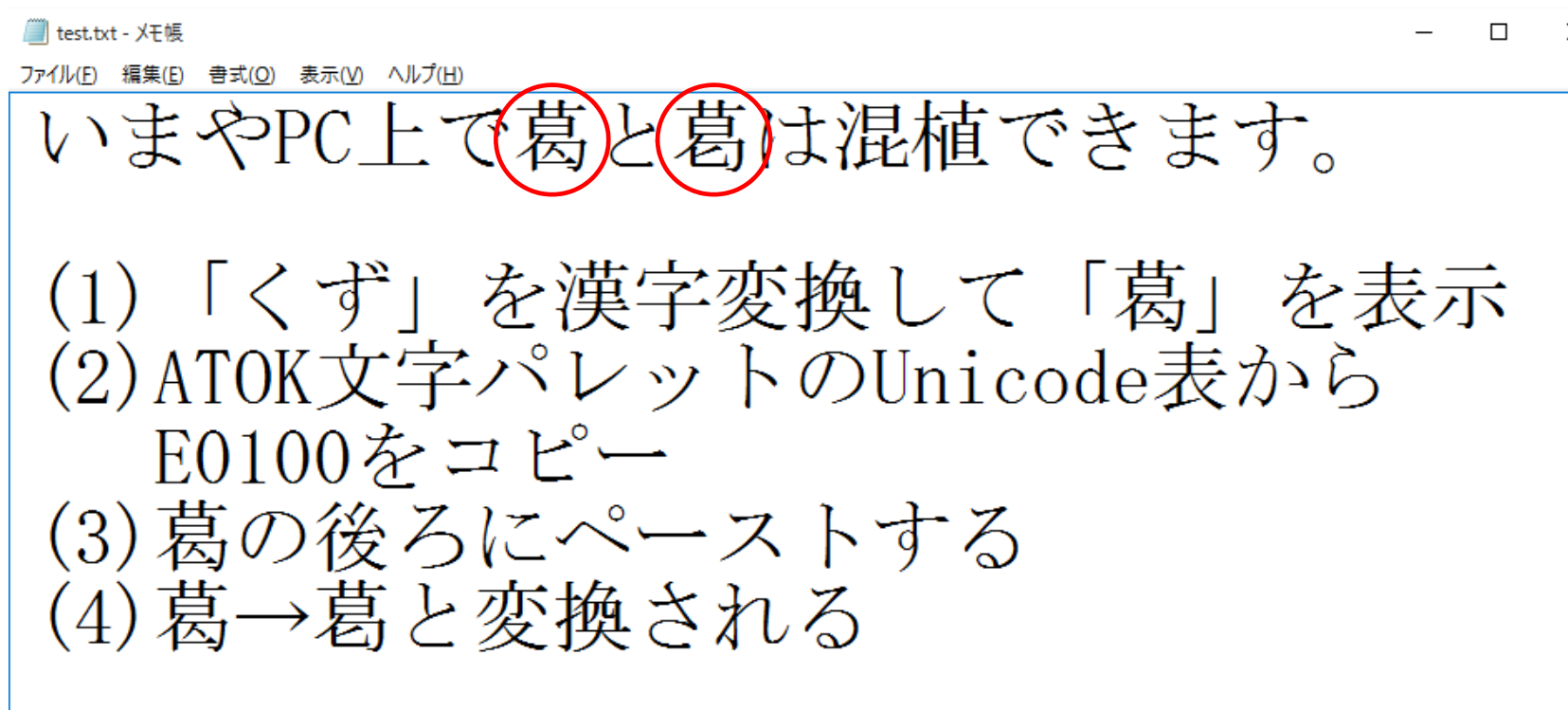
Moji-Joho : 「汎用電子・・・」の継続事業
「文字情報基盤構築に関する研究開発事業」

IVDの一部



IVSを使った表示例

環境：Windows10/メモ帳/MS明朝/ATOK



IVSを使った表示例

環境：Windows10/Excel2019/MS明朝

The screenshot shows the Excel interface with the formula bar displaying `=B4&UNICHAR(HEX2DEC(D4))`. The spreadsheet content is as follows:

	A	B	C	D	E	F
1						
2		04字形		IVS		90字形
3						
4		逢	+ E0100	→		逢
5						

IVS対応状況

■ OS

Windows7以降
MacOS10.6以降

■ アプリケーション

Adobe Reader 9以降
Adobe InDesign CS4以降
一太郎2014以降
Office2013以降

■ FEP

MS-IME
ATOK2014以降

■ フォント

MS明朝（JISX0213:2004変更字形のみ）
イワタOTF Pr6/Pr6Nシリーズ（対応コレクション「Adobe-Japan1」）
小塚明朝Pr6Nシリーズ（対応コレクション「Adobe-Japan1」）
IPAmj明朝（対応コレクション「Hanyo-Denshi」）

CJK互換漢字

■ CJK互換漢字とは

各国の漢字コードとUnicodeとで相互変換するために設けられた漢字領域

■ 本来は使用禁止！？

特定の規格との互換性を維持する目的以外で使用するはならないというのが原則

■ 互換漢字はなくなるかも

互換漢字に相当する統合漢字のVS (Variation Selectors) が追加標準された
(互換漢字と同じ文字が別のコードポイントに存在)

互換漢字はUnicodeの正規化プロセス(※)が走ると統合漢字に寄せられる
※文字の比較や検索をするためにOSが内部的に別のコードに置換する手法

文字情報基盤

経産省が主体となり各自治体が管理していた漢字を統合した電子政府用文字
(住基19563文字と戸籍55270文字を統合したもので計約6万字)

平成25年の閣議決定「世界最先端 IT 国家創造宣言」に
「文字の標準化・共通化に関しては、国際標準に適合した文字情報基盤を活用することを原則とする」
と記載されたことから注目されることになる。

IPAmj明朝としてフォント化され無償ダウンロード可能

2017年にすべての文字が符号化（Unicode化,ISO化）された。

2020年までに全国の自治体に普及することを目標にしている。
当面自治体主体で使用される予定だが民間で使用される可能性もある。

拡大するUnicode

最新のバージョンは12.1（2019年11月現在）で収容文字数は137,929文字
理論上は21億文字まで収容可能

中国の古代文字、少数民族文字などを第3面のCJK統合漢字拡張Gに
追加する計画もある

各国の言語はもちろん、近年では変体仮名、ARIB外字、絵文字、顔文字
なども収容しており内容は多岐に渡る

特に絵文字の多様化は著しく毎年のように追加されている
（髪の色とスタイル、肌色、性別、家族構成など）

多様化する文字

変体かな

阿 以 侍 意 悔 字

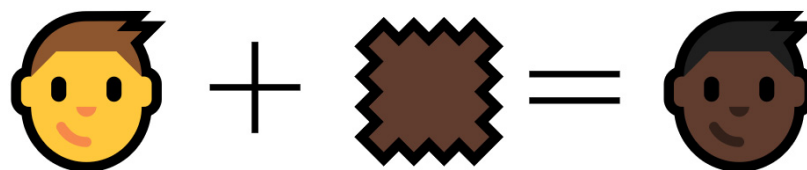
ARIB外字



絵文字



顔色



U-PRESSとは

U-PRESSとは

共同通信社が策定した15269文字の文字セット

目的

加盟社への配信用（2003年から配信開始）

新聞各社ではU-PRESSをベースにシステムメーカーやフォントメーカー、新聞社固有の文字を加えた文字で新聞製作している

採用規格・字種

JISX0201:1997

JISX0208:1997

JISX0212:1990（補助漢字）

JISX0213:2000

K-JIS固有非漢字

UCSの一部非漢字

注1) 2002年3月時点の上記規格の文字を採録

注2) 国語審議会による1022文字の表外漢字字体表（2000年）

注3) CO-77とJISとの字形差異はJIS優先

注4) JISX0212とJISX0213の重複文字はJISX0212の字形を優先

特徴的なU-PRESSのマッピング

2byte固定長でBMP（第0面）のみ使用

Unicode2面の文字をBMPの私用領域に移動させている

U-PRESSが策定された2002年当時はWindowsXPの時代
Unicode2面の文字は使用環境が整っていなかったため問題にはならなかった。

2018年現在、MS明朝はUnicode2面の文字を収容し文字変換も可能。

例えば「しかる」を漢字変換すると
叱 U+53F1
叱 U+20B9F
の両方が表示される



叱る

(MS-IMEの場合)

実際にあった問い合わせ

株式会社イワタ 水野様

お世話になっております。●●新聞の▲▲です。
フォントの表示で異常と思われる状況が発生しました。

明朝体、ゴシック体とも下記の文字が、他のフォントに置換されてしまいます。
「丈」「土」「汙」「碕」「脇」

ワード系のソフトや、Illustratorでも同様です。原因が分かりますでしょうか？

お忙しい中まことに恐縮ですが、よろしくお願い致します。

●●新聞社 営業局 広告センター

▲▲▲▲

回答

●●新聞 ▲▲様

株式会社イワタの水野です。
お世話になっております。

ご指摘いただいた文字はいずれもUnicodeの2面の文字であり、
一昔前のシフトJISのセットには入っていなかった文字です。

実はこれらの文字は最近のOS(WindowsVISTA以降)の
MS明朝などには入っているものの、新聞業界で標準と
なっているU-PRESSセットでは別の私的コードに
割り当てています。

理由としては共同通信社が新聞業界で標準的なU-PRESSを策定
した当時、Unicode2面を使用できるOSやアプリがあまり
無かったこともあり、Unicode2面の文字を共同外字として
Unicodeの私用領域に割り当てたことによります。

従って御社(に限らず新聞社のほとんど)の新聞編集の現場においても
上記の文字は何らかの形でコードを変換して使っているのが実態であり、
文字を巡る環境が過渡期であるが故の現象です。

以上、よろしくお願い致します。

	Unicode	U-PRESS
丈	2000B	EF32
土	2123D	EF33
汨	23C7F	EFB2
碯	2550E	EF40
脇	266B0	EFF4

U-PRESSの課題

1. UCS-4との非互換性

- ・ 共同外字エリアに配置されているUnicode2面の文字

→Unicodeとの互換をとる必要があるかも（2バイト固定長からの脱皮）

2. 拡張・変化するUnicodeへの対応

- ・ 新元号：来年5月に改元の際新元号の合字（新聞各社での対応が決定）
- ・ 変体仮名：おくやみ欄で使う社あり
- ・ IVSによる異体字流入の可能性
- ・ 外国語（簡体・繁体）
- ・ カラー絵文字（紙媒体以外のサイネージやWeb用途）
- ・ CJK互換漢字の憂慮

→規格化されているこれらの文字をいつまで外字として作る？

以上の課題に対し、システム側で対応するかフォント側で対応するかいずれも可能だが、**このような問題があることを念頭に置く**必要がある。

外字や画像はダメなのか？

文字コード（Unicode）化することは以下を可能にすること

- ・ データ交換
- ・ 検索／置換
- ・ 音声読み上げ
- ・ 自動翻訳

記者や読者から入稿があったり、記事をWebやサイネージ、放送に転用したり、あらゆる物がネットでつながりデータ交換する時代

できるだけ標準化されたコード（＝Unicode）を用いるのが望ましい

ご静聴ありがとうございました