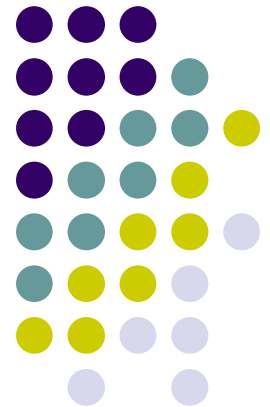


【マルチメディア文化論】

文字データの世界 (日本語文字)

2008年度 第6回

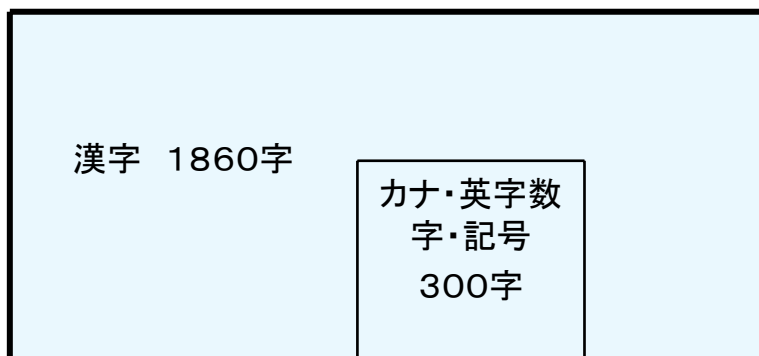
太田 信宏





1. コンピュータで漢字を扱うには

■ 漢字キーボード（漢字鍵盤）の時代



(横) × (縦)
60字 × 36字 = 2160字

■ 日本語ワープロの誕生

東芝製 JW-10 昭和53年(1978年)

[JW-1、JW-10の紹介](#)

<http://www.asahi-net.or.jp/~SD5S-SKMT/wpsokki10.htm>

[写真で見る歴史的なコンピュータ](#)

<http://www.watch.impress.co.jp/pc/docs/article/20010313/ipsj.htm>

■ かな漢字変換・ローマ字漢字変換の時代へ

AI変換 (Artificial Intelligence) → 変換精度の向上



2. 漢字の数

<日常使用する漢字>

学習漢字 → 1006字

常用漢字 → 1945字

人名漢字 → 488字

<大漢和辞典の収録漢字>

およそ10万字

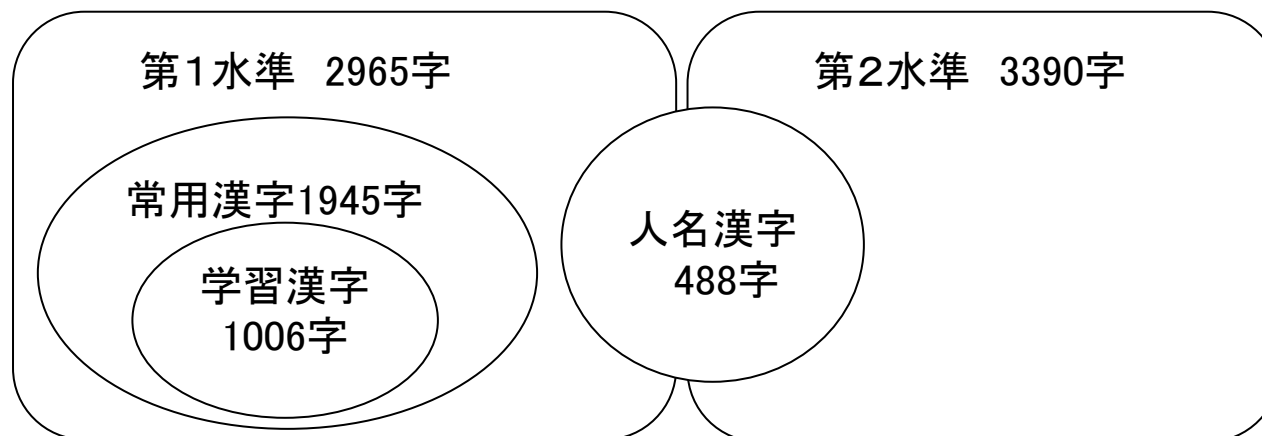
(多くの異体字あり)

劔, 劔, 劔, 劔, 劔, 劔など

<コンピュータの漢字>

◇第1水準 → 2965字

◇第2水準 → 3390字



[常用漢字の改訂記事](http://sankei.jp.msn.com/life/education/080513/edc0805131055003-n1.htm)

<http://sankei.jp.msn.com/life/education/080513/edc0805131055003-n1.htm>

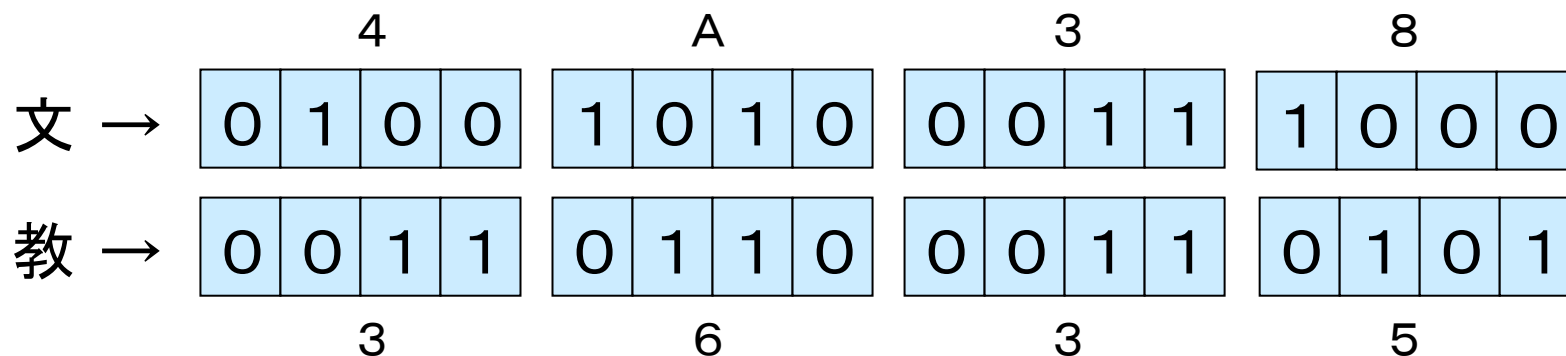


3. 全角文字のコード

半角 1 文字は 1 バイト = 8 ビット → 256 通り

全角 1 文字 = 2 バイト = 16 ビット → 何通り？

(JIS漢字コードの例)



第 1 水準と第 2 水準の合わせて約 6400 字は、
すべて異なる 0 と 1 の並びを持っている。

(文字を 4 桁のコード番号で識別)

文 → 4A38

教 → 3635



4. 欠字と外字

- ◇**欠字** コードが存在しない漢字のこと。
(コンピュータに登録されていない文字)
- ◇**外字** 欠字や特殊記号に対して、あとから任意に0と1(のコード)を割り当てた文字。

メーカー外字 ……パソコンに内蔵されている外字
(**機種依存文字**)

高 崎 青 朧 ① ② ③ I II III 絵文字、など多数ある。

ユーザー外字 ……利用者が独自に作成した外字

- ◇メールの外字使用はマナー違反。

使ってよい文字 → **JIS漢字コード** (第1水準+第2水準)



5. 文字コード例

	JIS漢字	シフトJIS	EUC	Unicode	種類
文	4A38	95B6	CAB8	6587	第1
教	3635	8BB3	B6B5	6559	第1
高	3962	8D82	B9E2	9AD8	第1
高	(967E)	(FBFC)	—	9AD9	外字
盆	4B5F	967E	CBDF	76C6	第1
桧	4930	95F4	C9B0	6867	第1
檜	5B58	9E77	DBD8	6A9C	第2
①	(2D21)	(8740)	—	2460	外字
I	(2D35)	(8754)	—	2160	外字



6. いろいろな文字コード

① JIS漢字コード（日本工業規格）

インターネットで利用可能な文字コード（第1水準＋第2水準）

② シフトJISコード

マイクロソフト社が日本のパソコン用に開発した文字コード

③ EUCコード（拡張UNIXコード）

主にサーバー（パソコンの上位機種）で利用される文字コード

↓ **新たな提案**

④ Unicode（ユニコード）

世界中の文字を1つのコード体系に収めるという壮大な構想。

英字、漢字（日本・中国・韓国・台湾）、ラテン、タイ、ハングル文字など

→ ①～③で表示できなかった多くの外字をカバー。



7. 文字コードの課題

① 欠字、外字の問題（JIS漢字、シフトJIS、EUC共通）

- ・過去の文献などをコンピュータに入力できない

② 略字体の代用ですむか？

- ・森鷗外
- ・高木さん→高木さん

③ Unicodeの課題

- ・各国の漢字を統合したことの問題 →（例）中国の「骨」という字
- ・異体字の扱い（コードセパレート問題）

文字コード情報に関するサイト

<http://www.shuiren.org/>

「立命館大学」山田崇仁先生

（文字コード入門）



8. 外字の調べ方

Windowsヘルプで「文字コード表」を開く

文字コード表

フォント(F): MS 明朝

①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮	⑯	⑰	⑱	⑳
I	II	III	IV	V	VI	VII	VIII	IX	X									
ミ	キ	ホ	ネ	グ	ト	ル	メ	ソ	フ	サ	ド	セ	ベ	シ	テ			
mm	cm	km	mg	kg	cc	m ²												
靴	”	、	No.	K.K.	Tel													
Ⓔ	Ⓕ	Ⓖ	Ⓗ	Ⓙ	Ⓚ	Ⓛ	Ⓜ	Ⓝ	Ⓞ	Ⓟ	Ⓠ	Ⓡ	Ⓢ	Ⓣ	Ⓤ	Ⓥ	Ⓦ	Ⓧ
≡	≡	∫	∫	∑	√	∟	∠	∠	∠	∠	∠	∠	∠	∠	∠	∠	∠	∠
繡	襲	鍔	鍔	醜	倍	焔	昱	精	鋸	剪	乞	任	佞	仔	但	佻	佻	佻
佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	徹	兇	燿	宜	冷	風	弱	弱
尢	尢	尢	尢	尢	尢	尢	尢	尢	尢	尢	尢	尢	尢	尢	尢	尢	尢	尢

コピーする文字(A):

詳細表示(V)

文字セット(T): Windows: 日本語

グループ(G): シフト JIS カテゴリによる入力

検索する文字の名前(E):

U+5F45 (0xFA67): CJK Ideograph

グループ

シフト JIS カテゴリによる入力

- 半角文字
- 記号、桁およびラテン
- ひらがなとカタカナ
- ギリシャ文字
- キリル文字
- 特殊文字**
- JIS 第 1 水準
- JIS 第 2 水準