

【マルチメディア文化論】

文字データの世界（日本語文字）

2004年度 第6回

太田 信宏



1. 漢字をコンピュータで扱うには

漢字キーボード（タブレット）の時代

漢字 540字	漢字 540字	漢字 240字	漢字 540字
		かな記号 など 300字	

(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)

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

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

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	284字

第1水準

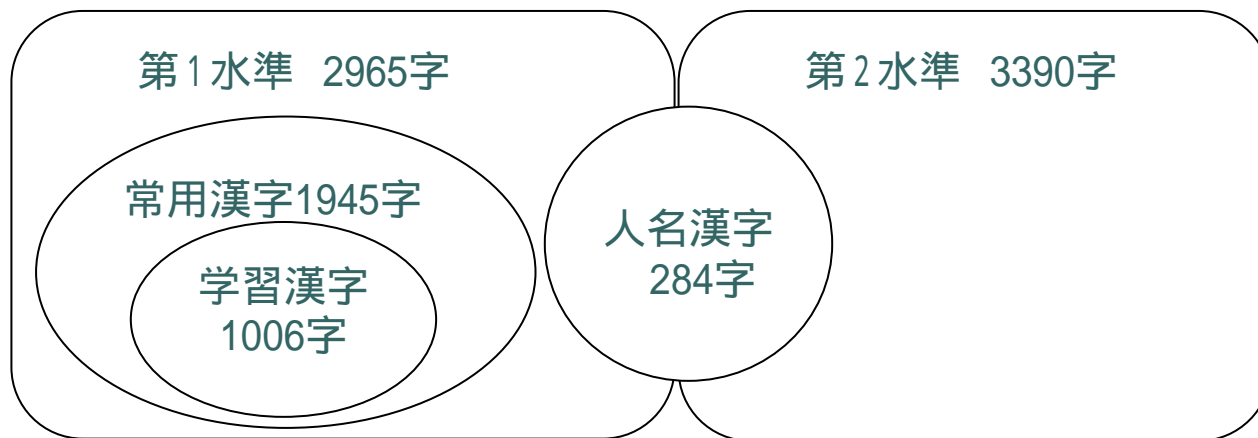
第2水準

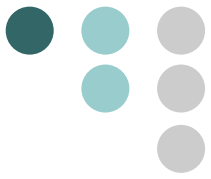
< 漢和辞書の収録漢字 >

大漢和辞典 およそ10万字
剣, 劍, 劔, 劔, 劔, 劔など

2965字

3390字





3. 全角文字のコード

全角 1 文字 = 2 バイト = 16 ビット
(半角 1 文字は 1 バイト)

(JIS漢字コードの例)

文	0	1	0	0	1	0	1	0	0	0	1	1	1	1	0	0	0	0
教	0	0	1	1	0	1	1	0	0	0	1	1	1	1	0	1	0	1

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。

(インターネットのルール)

メールやホームページには、**JIS漢字 (第 1 水準 + 第 2 水準)** の文字だけを使用することができる。



4 . 欠字と外字

欠字 コードが存在しない（コンピュータに登録されていない）漢字のこと。

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

高 崎 青 彌

……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

インターネットの世界では**どちらの外字も使用禁止。**

5. 外字の調べ方

書籍の活用 「JIS漢字辞典」他
(文字コードに関する書籍は多数あり)

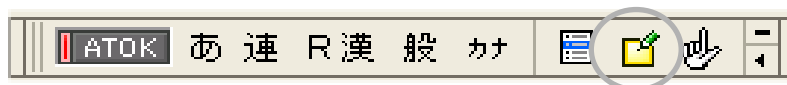
日本語入力ツールバーで調べる

MS-IMEパッド

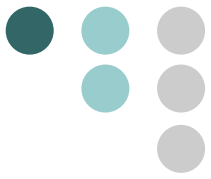


「ヘルプ」 - 「目次とキーワード」 - 「入力しにくい文字の入力」
IBM拡張文字コード、NEC選定文字コードなど

ATOK文字パレット



「和文コード表」 第1水準、第2水準、外字などが収録



6. 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
	(2D21)	(8740)	-	2460	外字
	(2D35)	(8754)	-	2160	外字



7. いろいろな文字コード

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)

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

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

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

~ で表示できなかった多くの外字をカバー。



8 . 文字コードの課題

漢字が足りない (JIS漢字、シフトJIS、EUC共通)

- ・欠字、外字の問題
- ・過去の文献などをコンピュータに入力できない

略字体の代用ですむか

- ・森鷗外
- ・高木さん 高木さん

Unicodeの問題

- ・コンピュータメーカー主導で提案
- ・各国の漢字を統合したことの問題 (例) 中国の「骨」という字

文字コード情報に関するサイト (立命館大学 山田崇仁先生)

<http://www.ritsumei.ac.jp/kic/~tyv07679/chuden/teach/code/index-j.html>