

外字・異体字の バリアフリーを目指して 「こだわりの文字」をどうするか

NPO文字文化協會
インデックスフォント研究会
黒田信二郎

TIBF2014 **出版梓会** 人文社会科学イベントスペース
2014.07.03

日本の漢字1600年の歴史

■日本の漢字1600年の歴史 「漢字は日本語である」

漢字をルーツとする多様な文字表現の自由が確保される出版環境が求められてきた

■外字・異体字とは

・外字

- ①規範・規格(常用漢字表やJIS漢字コード、Unicode)の「外にある」
- ②機器(入力・表示・印刷)に実装されているフォントセットの「外にある」・・・文字コードがあるからといっても使えない！

・異体字

コンピュータ以前の文字(手書き文字・印刷活字)は多種多様なカオス(漢字研究の領域)であった・・・

現場では、ゲタ文字・文字化けの問題として現れる！

多様な漢字の世界



国字・おとど
066147



国字・いずな
089065



国字・きょうだい
065036



国字・たばか(る)
065415



国字・じじ
065413



国字・たか
066096



国字・げそ
065987



国字・このしろ
065947



国字・いるか
065980



国字・ふぶき
065041



国字・あられ
065886



国字・ざんぶと
065229



古字・シャ
024714



古字・サイ
018923



同字・直
212546



同字・祭
074376



名称文字・瀉
201894



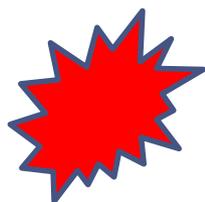
名称文字・堂
210248

(6ケタの番号はインデックスフォント (文字鏡) 番号)

印刷・出版の文字のデジタル化

あらゆる工程にデジタル化の波

- ・文字コードの世界⇒標準化、効率化、少数化
文字の抽象的概念を符号化したもの
包摂、統合がある、字形は例示である
- ・印刷・出版の現場⇒多様な表現を追求
一意に指定された図形を、差異を区別して送り
受けること(1対1の関係)が求められる



文字コードでは文字図形がユニークにならない問題が
混乱につながった

文字コードの役割

- コンピュータで扱うための「標準セット」となる基盤
異なった環境でも汎用的に使える**文字集合として不可欠の存在**
処理の効率化のため少数化を視野にいれる(統合/包摂/縮退)
- 印刷・出版の現場処理に必要な図形とは (**ロングテール問題**)
「滅多に使われない字」の分布と文字コード(実装フォントセット)対応を
どう考えるか
- 日本語の印刷・出版では**JIS X 0213:2004**がひとつの境界と考えられる
JIS漢字第1水準～第4水準の10,050字
一般的な漢和辞典の収載文字数も10,000字前後
- 国際対応にはUnicodeが必須**
システムの内部処理コードとしての汎用性

「印刷」と「表示」

- 印刷書籍 企画⇒執筆⇒編集⇒制作⇒印刷⇒流通
印刷ができればとりあえず解決
- 電子書籍 企画⇒執筆⇒編集⇒制作⇒流通⇒表示
読書端末のフォント環境に依存

対応方法

- 印刷・表示可能な文字に置き換える(置き換えの判断の正否)
元の字形に戻せなくなる

- 専用フォントを作成してフォントを埋込む(外字エリアを利用)
- 外字を画像で貼り込む
印刷・表示は可能だが、あくまでローカル処理のため、
情報交換・共有、文字検索ができない

「こだわりの文字」とは

こだわる(区別する)必要性はTPOによる

- ・日常的な生活や学校教育に必要な範囲(常用漢字・教育漢字)
- ・住所・氏名の表記(行政やDM・顧客管理)に必要な範囲(電子政府文字)

■学術歴史文化や創作的表現に必要な範囲(出版のニーズ)

◆識別の粒度(とめ、はね、はらいはデザイン差だが・・・)

「こだわりで区別する必要」の判断

区別する必要がある例(文字図形の差異に意味があるもの)

- ・作品・作家研究(文字の使い方に作品・作家の個性がある)
- ・歴史的変遷(文字が時代の違いを示す)
- ・固有名詞(固有の違いを字形が示す)

◆明らかな誤字・誤植(無駄な外字を作らない・・・)

◆印刷・表示のフォント(サイズもいろいろ、デザインもさまざま・・・)

「版面レイアウト」「フォント」の選択

「基準」は「明朝体」だが書体・デザインはフォントベンダーの領域

印刷・出版の文字バリエーション

・文字は抽象概念

山 山 山

・国字・創作漢字

峠 鮎 庖

・識別の粒度

邊 邊 邊

(以上は文字鏡明朝SVG画像フォントで表示。色付き文字は規格内例字図形)

SVG: Scalable Vector Graphics

印刷・出版のフォントバリエーション

・大きさ

鷹 鷹 鷹 鷹 鷹 鷹 鷹 鷹 鷹

・書体・デザイン

美 美 美 美 美 美 美

表現の多様性はフォントの大きさ、書体・デザインの選択にもある
⇒ 版面レイアウト

なぜ「文字図形番号」か

文字コードでは扱えない文字のピンポイント対処方法

- ・文字コードにない
- ・文字コードでは一意に図形が決まらない(恐れがある)文字

※Unicodeは統合漢字の拡張と異体字枝番(IVS)で対応

IVS: Ideographic Variation Sequence/Selector

「外字」「異体字」など文字コードでは扱えない文字で、区別する必要がある「図形」に

- 固有の名前(文字図形番号)をつけ、
- その番号を共有化することで、
- 次世代に伝えるべき、精確な「文字情報」を継承する
(もともと漢和辞典などは「字典番号」を用いている)

文字図形番号に関連する規格・標準(文字鏡研究会の成果18万字を参照)

◎国際規格:ISO/IEC 10036(JIS X 4165)

Information technology – Font information interchange – Procedures for registration of font-related identifiers

◎情報処理学会試行標準:IPSJ-TS 0002:2014「文字図形識別情報」

漢字の図形的違いを識別する(同定)規則

⇒インターネットで文字図形を公開、図形の確認ができる環境

「外字・異体字」の制作手法

- こだわりの文字を求める著者には、執筆入稿時から該当文字に「**文字図形番号**」を指定してもらう

(例) 森鷗外は毛筵を「搗 (&M057674;) 」と書いた。

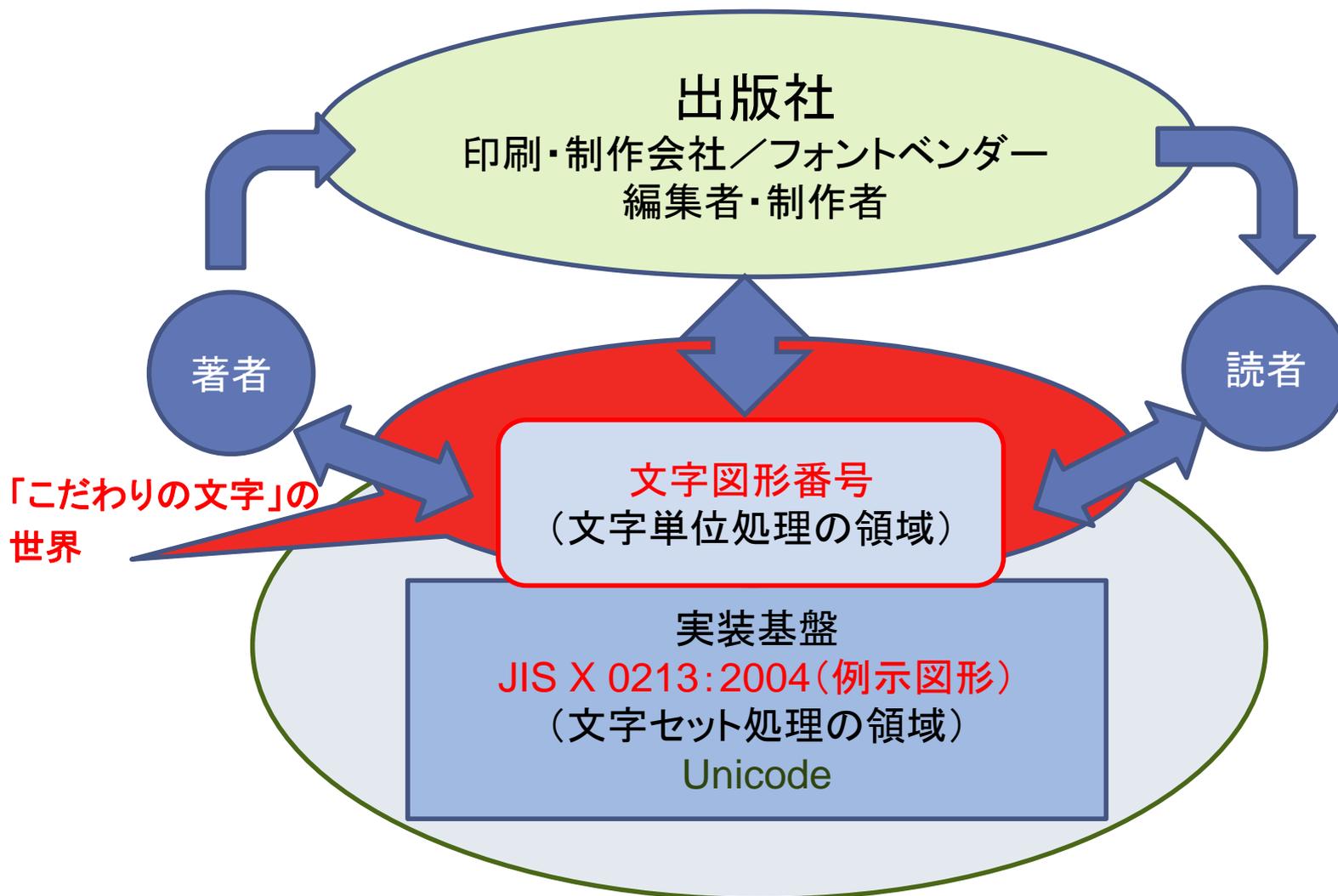
「雑司ヶ谷 鬼 (&M076218;) 子母神」は安産子育ての神様として知られる。

(外字は文字鏡明朝TTFフォントで表示。文章はリュウミンProL-KLを使用)

- 制作するデータで、現状の実装環境では印刷・表示ができない恐れがある文字は、「**文字図形番号**」を付与して管理しておく
- ローカル外字対応や「画像貼り込み」でも、何らかの形で「**文字図形番号**」を保存しておく

文書形式が変更されても「文字図形番号」で
図形との1対1対応が保証される

文字図形共有環境のイメージ



共有環境の必要性と課題

「文字図形番号」とのテーブルを介在させることで
一意の文字図形がデジタル環境で必ず継承される
基盤となる共有環境を作ろう！

100年後のコンピュータ技術(アプリケーション)が
どうなっているか誰もわからない！

(課題)

文字単位処理の手法改善

- ・WEBフォントの技術進展
- ・図形番号による入力や検索の方法の改善
- ・フォント(セット)書体・デザインのバリエーションとリソース指定
- ・新たに出現する、区別する必要がある文字図形の登録

ご清聴ありがとうございました

「外字・異体字のバリアフリーを目指して 漢字研究7年の軌跡」Kinoppy版

電子書籍版制作の手法について

- ・本書の制作にあたっては、「外字・異体字のバリアフリーを目指す」観点から、現段階で印刷書籍並びに電子書籍のあらゆる環境・工程で共有可能な基本キャラクターセットがShift_JISであることを前提とした。
- ・それ以外の文字(甲骨文字、梵字を含む)については、インデックスフォント番号(文字鏡番号)を指定し、印刷書籍では文字鏡明朝TTFフォントを利用し、電子書籍では文字鏡SVG画像フォントの文字埋め込み処理がされている。
- ・本文中で文字に付してある6ケタの番号は、インデックスフォント番号を示し、ISO/IEC10036規格サイトRegistry(<http://10036ra.org/glyph-index.html>) Mojikyo欄[10XXXXXX]で字形が確認でき、グリフ画像のコピーも可能である。
- ・なお、印刷書籍版下作成はDTPソフト「EDICOLOR10」(キャノンITソリューションズ)を用いた。また、電子書籍EPUB3データは印刷書籍版データをもとに有限会社ワイズネットで作成した。

電子書籍版には「縦書きページ」のサンプルも収載！