Vol.2011-DD-81 No.6
2011/7/22

電子書籍における外字・異体字問題に関する一考察

秋元良仁[†]　高田智和[††]　小林龍生[†††]

電子書籍に代表されるデジタルコンテンツの利用環境においては，多様な文字表現が求められている．しかしながら，制作環境や表示デバイスの違いによって表示可能な文字は異なるため，外字や異体字等，多様な文字表現は困難となっている．本稿では，経済産業省による外字・異体字の整備事業を中心に，電子書籍における外字・異体字問題を示し，その解決手法について考察する．

A Study on External Characters and Ideographic Variant Characters issue on E-Book

Ryoji Akimoto[†] Tomokazu Takada[††] and Tatsuo Kobayashi[†††]

Recently, information technology has progressed. Then, the amount of digital contents that can be used increases. In such a situation, it is necessary that the environment of digital contents require various character representations. In this paper, we summarize the current state of digital characters and problem. Based on it, we introduce "External characters / ideographic variant characters solution project" by Ministry of Economy, Trade and Industry. And we propose the concept for how to solve the problem of external characters / ideographic variant characters. And then, we describe the design of environment as readily available to such characters.

1. はじめに

デジタル・ネットワーク化された環境において，日本語の特徴である多様な漢字表現は，端末機器上で電子的に表示し，かつ広く伝播する流通システム上で取り扱うために規格の平準化・限定化が要求される．しかしながら，著作者・出版社においては自らの意図を正確に表現したい，学術上の正確さを表現したい等の要求があり，また読者の中にも電子的環境において漢字表現の多様性を求める者もいる．

このような現状に対し，経済産業省では出版物の利活用促進のための外字・異体字利用環境について，外字の収集・整理方法，文字図形の共通基盤の運営方法，利用端末での外字実装方法，電子的環境での円滑な外字・異体字の配信方法等を包括的に検討するプロジェクトを実施している．

本稿では，プロジェクトの方向性を検討する上で基礎的な資料として用いられた凸版印刷株式会社の漢字出現頻度数調査の概要を示すとともに，プロジェクトの概要とその解決手法を示し，電子的な環境における外字・異体字問題について考察を加える．

2. 漢字出現頻度数調査

2.1 概要

文化審議会国語分科会では，2005 年 3 月の文部科学大臣の諮問「情報化時代に対応した漢字政策の在り方について」検討するため，同年 9 月より漢字小委員会を設けて審議を行った．審議用の基礎資料として，凸版印刷は文化庁に対し，凸版印刷が保有する書籍に関するデータを用いて漢字の使用頻度の実態調査を報告している．調査は「漢字出現頻度数調査」（文化庁文化部国語課，1997 年 11 月），「漢字出現頻度数調査（2）」（同，2000 年 3 月）を受け，「漢字出現頻度数調査（3）」[1)]（同，2007 年 3 月）としてまとめられている．なお，「漢字出現頻度数調査（3）」では，凸版印刷が 2004 年から 2006 年に作成した組版データを用いている．

2.2 調査対象書籍

調査対象の書籍は「辞典類」「単行本」「週刊誌」「月刊誌」「教科書」の 5 分野とし，分野毎にデータ量のバランスを損なうことがないよう，調査対象漢字数の比率を「教科書」を除く 4 分野において「辞典類」「単行本」「週刊誌」「月刊誌」の順に「1：3：

† 凸版印刷株式会社
TOPPAN PRINTING CO., LTD.
†† 国立国語研究所
National Institute for Japanese Language and Linguistics
††† 有限会社スコレックス
Scholex co., ltd.

1：1」の程度になるように配慮している．表 1 に教科書を除くサンプリング書籍の内訳を示す．

表 1　サンプリング書籍の内訳

分野	書籍冊数	出現文字数		出現漢字数	
単行本	540	88,189,211	53.9%	24,858,027	51.9%
月刊誌	120	32,971,129	20.2%	9,560,173	19.9%
週刊誌	150	23,477,267	14.4%	7,688,151	16.0%
辞典・事典	12	18,849,349	11.5%	5,818,082	12.1%
合計	822	163,486,956	100.0%	47,924,433	100.0%

2.3 外字・異体字の出現頻度

漢字出現頻度数調査は，一般の人々の文字生活において大きな役割を果たしている書籍等の漢字使用の実態を明らかにすることを目的としており，漢字小委員会では調査に基づき，出現頻度数の高い漢字に着目して「漢字使用の目安としての漢字表」の整備を検討している．

他方，本研究では，国内の一般的な書籍においてどの程度符号化されていない外字・異体字が存在するのか，その出現頻度が低い漢字に着目する．これらの漢字に対して解決策を提案することで，3 章以降で説明する外字・異体字が容易に利用できる環境の整備を行うことを目的としている．

漢字出現頻度数調査に基づき，表 2 にどの程度異体字（IVS の候補となりうる文字）が出現しているのかを示す．表 2 の出現漢字数は表 1 の出現漢字数と同値である．また，出現漢字数はサンプリング書籍において同一漢字が複数回出現した場合，各々1 文字とカウントしているのに対し，出現字形数は同一漢字が複数回出現した場合はそれらをまとめて 1 文字としてカウントしている．

表 2　IVS 候補文字

分類		出現漢字数		出現字形数	
2-1	正字	47,704,927	99.5%	7,626	88.9%
2-2	異体字	219,506	0.5%	950	11.1%
合計		47,924,433	100.0%	8,576	100.0%

表 3 に符号化されていない漢字（SVG 等画像化の候補となりうる文字）がどの程度出現しているのかを示す．表 3 も出現漢字数は表 1 の出現漢字数と同値である．表 3 では，サンプリング漢字に対し，Unicode および CID（Adobe Systems が定める文字集合仕様．文字ごとに一意の数字番号が割り当てられる．文字集合を示す書式は「登録者－配列（－追補番号)」であり，日本の場合，Adobe-Japan1-6 が最新となる.）が割り当てられている漢字，CID のみが割り当てられている（Unicode なし）漢字，Unicode も CID も割り当てられていない漢字に分類している．Unicode および CID が割り当てられている漢字は，更に JIS X 0208 に該当する漢字とそれ以外の漢字に分類している．

表 3　SVG 候補文字

分類			出現漢字数		出現字形数	
3-1-1	UNICODE / CID	JIS X 0208	47,542,535	99.2%	5,774	67.3%
3-1-2		JIS X 0208 以外	70,049	0.1%	1,426	16.6%
3-2	CID のみ		140,028	0.3%	393	4.6%
3-3	上記以外		171,821	0.4%	983	11.5%
合計			47,924,433	100.0%	8,576	100.0%

表 2 および表 3 から，凸版印刷の調査に基づいて考えると，日本の出版物における漢字表現は約 99.6%（表 3 の 3-1-1～3-2「出現漢字数」の合計）が国際規格と整合性のある符号化方式で表現可能となり，それ以外の約 0.4%がユニークな名前を持つ図形（SVG 等の画像化候補）で表現することとなる．

3. 外字・異体字が容易に利用できる環境の整備プロジェクト

総務省，文部科学省，経済産業省は共同の懇談会を開催し（2010 年 3 月～6 月)，「デジタル・ネットワーク社会における出版物の円滑かつ安定的な生産と流通による知の拡大再生産の実現」を目指すための一方策として，経済産業省を主担当とした「外字・異体字が容易に利用できる環境の整備」プロジェクト[2)]を発足させた．

プロジェクトは凸版印刷株式会社を事務局とし，日本文藝家協会の三田誠広氏を座長とする有識者や業界関係者による専門家委員会を設置，以下の 4 点について調査および提案の検討が行われた（2011 年 1 月～3 月)．

＜調査分析項目＞

(1) 印刷・出版業界の「外字」の現状調査

印刷会社における制作ワークフロー（CTS 方式と DTP 方式）において，外字制作をどのように対応しているのか現状調査を行う．また，デジタルコンテンツ配信におけ

る外字・異体字の取り扱いについて，対応状況・運用ルール・課題等の現状調査を行う．

(2) これまでの「外字・異体字」問題に対する動向調査

これまで国内を中心に実施されてきた大規模文字集合プロジェクトに対してヒアリングを実施し，各プロジェクトの目的，概要，実績，課題等を調査する．ヒアリング対象は以下の 7 プロジェクトである．

1. 文字鏡研究会 [3)]
2. インデックスフォント研究会 [4)]
3. GT プロジェクト（TRON プロジェクト）[5)]
4. CHISE プロジェクト [6)]
5. 漢字データベース [7)]
6. グリフウィキ [8)]
7. 文字情報基盤構築事業 [9)]

(3) 電子出版（日本語テキストのデジタル化）における文字に関する問題点調査

（1）外字の現状調査および（2）これまでの大規模文字集合プロジェクト動向調査を踏まえ，電子出版における文字の取り扱いに関する問題点の整理を行う．

(4) 書籍等のデジタル化に伴う「外字・異体字」問題解決策の提案

問題点の整理を踏まえ，「外字・異体字」問題に対して適切な方向性を示し，次年度以降期待される実証実験の具体的な実施内容・方法・課題について検討を行う．

本稿では，以下 4 章以降，調査・提案の概要を示す．

4. 印刷・出版業界の「外字」の現状調査

4.1 印刷会社における外字・異体字対応フロー

CTS（Computerized Typesetting System，コンピュータを利用した写植組版システム）および DTP（Desktop Publishing，デスクトップ出版）別に印刷会社にヒアリングを実施し，印刷会社における外字・異体字対応の実態を調査した．

4.1.1 CTS における外字・異体字対応フロー

図 1 に印刷会社における CTS での外字処理作業プロセスを示す. 外字制作は, 正字・旧字を共に利用したい書籍の制作時に発生する．例えば，古書やその解説書，異体字の例示を多用した辞典等において多く発生する．外字や異体字に関しては，印刷会社内で独自開発した文字検索ツールで内部コードに基づき管理しているため，印刷会社内で独自に制作した外字も含めて統一的に管理することができる．その対象範囲は DTP 関連システム・旧ホスト系システムのデータも含めて対象となっている．

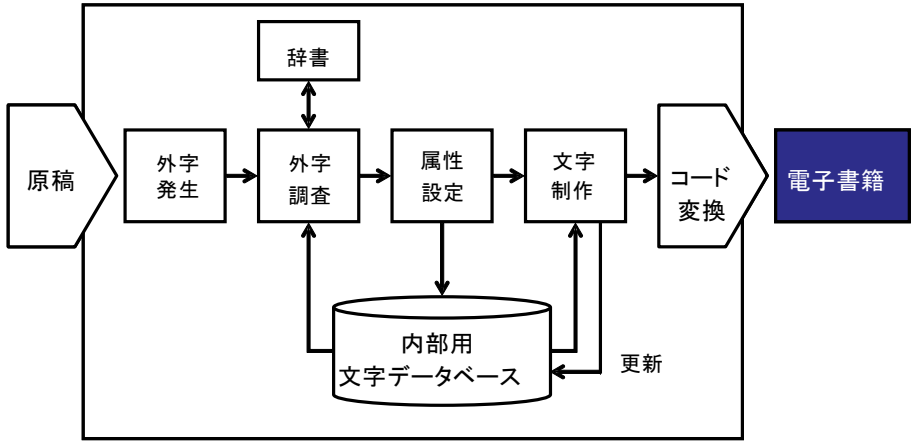

図 1　印刷会社における CTS 外字作業プロセス

4.1.2 DTP における外字・異体字対応フロー

DTP に関しては，出版社毎（あるいは編集プロダクション毎）に個別管理を実施している．そのため，印刷会社内で共通した文字コード管理は行われていない．個別管理で行われる外字の取り扱いは大きく分けて以下の 2 通りとなる．

（1）印刷会社内で作字する
（2）出版社等より外字ファイルとして供給を受ける

現状，上記（1）および（2）の対応は DTP ソフトウェア Adobe InDesign（文字集合としては Adobe-Japan 1-5 および Adobe-Japan1-6）でほぼまかなえており，CTS に見られる体系立てた管理は行われていない．

4.2 デジタルコンテンツ配信における外字・異体字の取り扱い

デジタルコンテンツ配信における外字・異体字の取り扱い状況を調査するため，デ

ジタル配信事業者にヒアリングを実施し，対応状況，運用ルール，課題等の調査を行った．

4.2.1 デジタルコンテンツ配信事業者の外字・異体字対応状況

デジタルコンテンツ配信における主な外字・異体字の取り扱いは，以下の 2 通りとなる．

（1）画像化
（2）JIS 第一水準・第二水準内の文字へ置き換え

同一の書籍タイトル内で混在して利用される場合や，特定の文字セットで用意可能な外字フォントにマッピングする場合等，例外処理については出版社と相談の上，その都度柔軟な対応を行っている．

4.2.2 運用ルール

外字・異体字への対応は，概ね以下に示すルールに基づき運用されている．

（1）出版社に確認
まずはコンテンツホルダーに対し，外字・異体字を画像化するか，あるいは JIS 内の文字に置き換えるのか方針を確認する
（2）画像化ルールの適用
画像化する場合，画像サイズとフォーマットに関する配信会社のルールを説明する．その上で例えば制作会社が保有するフォントに基づき表示用画像を制作する

4.2.3 課題

画像化する場合，端末毎のグラフィック特性（解像度や階調等）の違いによる表示フォントと外字画像との見栄えの違い，あるいは，表示フォントと外字画像の書体の違い（明朝かゴシックか等）による文字の不均一が生じる場合がある．
さらに，ユーザ操作によりアプリケーション側で表示フォントが切り替えられる可能性があり，その場合もやはり表示に不自然さが生じる場合がある．
また，将来的にフォントへの置き換え等が行われる場合，ファイル名の対応を取る必要があり，制作時の負荷が高くなることが予想される。

5. これまでの「外字・異体字」問題に対する動向調査

現在までに国内で行われてきた各種大規模文字集合プロジェクトに対してヒアリングを実施し、各プロジェクトの目的、概要、実績、課題等を整理した。以下にヒアリング内容の概要を示す。

表 4　文字鏡研究会ヒアリング内容

目的	・漢字とこれに属する文字、諸国の文字、かつて文化を支えていたが歴史に埋没している文字を利用可能にする調査研究 ・文字番号の採番とフォントの配布
概要	・1997 年 4 月に研究会発足 ・UCS を中心とした CJKV 文字、梵字、甲骨文字、西夏文字、非漢字等、合計約 16 万文字をカバーし、6 桁の独自文字番号を採番 ・非営利学術用途の会員に対して、無償でフォント利用や文字の作成申請が可能 ・ビジネス用途では、パッケージソフト「今昔文字鏡」が利用可能 ・書体は字形例示書体で明朝のみ
実績	・大蔵省印刷局「官報デジタル化」、国立公文書館 ・学術調査・資料作成 ・大学教育
課題	・契丹文字等、更なる歴史的な文字収集とフォント化 ・取り組みへの公的な支援

表 5　インデックスフォント研究会ヒアリング内容

目的	・コード表にない漢字等へユニークな文字番号付与を行い、対応する基準フォントの作成と文字属性情報付与を行い整備 ・新聞、出版、印刷、ビジネスフォーム、官公庁業務等の外字を含むテキストデータの汎用性を確保
概要	・文字鏡研究会の成果（約 16 万文字）を活用 ・業界が抱える文字問題の解決策の検討 - 文字入力、検索方法の検討 - 規格文字コードとの対応テーブルの検討 - 文字作成、登録、属性付与方法の検討 - 文字の同一性に関するルールの検討 等 ・講演会等による普及啓蒙、技術・標準化動向の把握
実績	・新聞、出版、印刷、ビジネスフォーム等の製作行程での字形判定
課題	・研究成果の実ビジネスへの展開 ・取り組みへの公的な支援

表 6　GT プロジェクト（TRON プロジェクト）ヒアリング内容

目的	・ユビキタス社会で、誰でも扱える TRON 多国語言語環境を実現させる
概要	・TRON は、さまざま言語を包含する文字セットを基盤として、その上位に文字入力・文字属性データベース等のアプリケーション層を持つトータルなアーキテクチャ（言語混在の状況に強い） ・GT 明朝 - TRON 多言語環境における漢字面の一部 - 諸橋大漢和をベースに約 6 万強の独自文字コード、例示字形を整備（現在は拡張されて約 10 万文字を収録） - グリフは TrueType フォント及びビットマップで利用可能 ・T 書体 - GT 明朝に含まれない中国漢字（漢籍、宋、明時代）を追加 - GT 明朝と併せて、約 13 万文字をカバー
実績	・図書館システム ・自治体システム等
課題	・T 書体は歴史学的な観点からの検証が困難

表 7　CHISE プロジェクトヒアリング内容

目的	・文字コードを使わないで文字処理が行える状況を確立させる ・符号化文字集合に含まれない文字も、区別無く容易に使えるようにする
概要	・文字を扱うメタ・システム（1999 年スタート） ・各文字に対し、字形（IDS）、部首、画数、文字コードへのリンク等と、それらの関係性をメタデータベースとして蓄積 ・諸橋大漢和、GT 明朝、全ユニコード等、約 28 万文字を構築 ・例示字形等のグリフそのものは保有していない ・CHISE Wiki として、漢字検索とメタデータ登録を Web サービスとして公開。グリフウィキとも連動
実績	・東洋学文献類目データベース化、組版及び検索システム ・グリフウィキ（メタデータ提供） ・CHISE IDS 漢字検索
課題	・メタデータの精度アップ（漢語的意味の追加、アクセシビリティ対応等）

表 8　漢字データベースヒアリング内容

目的	・検索等により、UCS（CJK 統合漢字）を扱い易くし、その利用を促進させる ・漢字の関係性の明確化
概要	・2003 年スタート ・漢字に関する周辺情報を整備 ・諸橋大漢和、仏典、情報処理学会試行標準規則等をカバー ・漢字辞書、字形（IDS）、異体字の三つのデータベースで構成
実績	・グリフウィキ ・学術資料作成、辞書として利用
課題	・データ活用手法の啓蒙

表 9　漢字データベースヒアリング内容

目的	・文字の“青天井問題” に対するソフトウェアによる解決
概要	・2007 年 10 月公開 ・ウィキペディアのように誰でもグリフ作成・登録・利用が可能 ・UCS 約 75,000 文字をカバー、漢字データベースを活用 ・大手フォントベンダーが取り組まないような、ニッチなニーズが当面のターゲット ・TrueType、SVG、PNG 形式で出力して使うことができる
実績	・符号化されていない文字を含む文書作成（学術利用） ・Web ページでの利用（学術利用）
課題	・利用拡大と認知度アップ ・プリント出力とのシームレスなフローの確立

表 10　文字情報基盤構築事業ヒアリング内容

目的	・行政処理の合理化（行政システムの構築、運用、保守に伴う氏名表記に関わる実務の利便性向上）
概要	・戸籍統一文字と住基統一文字を中心に、ISO/IEC 10646 や JIS 漢字コード等の漢字関連情報を整理統合した漢字情報テーブル及びこのテーブルに対応したフォント（或は漢字図形）から構成される ・新漢字情報テーブル 戸籍統一文字､住基統一文字と国際符号化文字集合の対応関係や各種属性情報等を収録し、漢字の異同確認、同定、交換用テーブル作成に利用できるテーブル ・IPA フォント 新漢字情報テーブルに対応した OpenType フォント。IPAex 明朝を中核にしてできている。また、IPA フォントライセンスに基づき、無償で利用が可能
実績	・IPAex 明朝フォントの提供 ・IPA 文字検索システムの提供
課題	・電子政府における利活用方法の検討 ・継続的な維持／運用体制の検討 ・文字情報一覧表の継続的整備

6. 電子出版における文字問題とその解決策

6.1 制作と利用

電子書籍に代表される電子的な環境で出版物の文字問題を取り扱う場合，出版物の制作（作り手側)，出版物の利用（読者側）という工程別に分けて考える必要がある.

これまでにも外字や異体字を含む文字に関する問題の解決に向けて多くの議論がなされてきたが，この視点が共有できていない議論では，例えば「読者側に全ての外字･異体字をカバーし得る膨大なフォントセットが必要だ」といった誤解を生じさせ，解決の糸口が見えない議論に陥りやすかったと言える.

表 11 に工程別の特徴と問題点を示す. 作り手側は更に執筆工程と編集工程に分類している．執筆工程では，コンテンツの創造活動という特徴に対し，執筆者が編集工程に対して直接外字や異体字等の指示を伝達できないという問題がある. 編集工程では，日本語の性質上，文字の出現頻度とは無関係に膨大な表現用途の字形が存在するという特徴に対し，どの文字が外字・異体字であり，どの文字が規格化された文字であるのかその判定基準が不明確である，データ形式が不統一であるために流通システム上の情報互換性が乏しいという問題がある．また，読者側では，閲覧端末やアプリケーションによって対応する符号化文字集合が変わり，それに応じて表示形式も変化するという特徴に対し，外字や異体字が正確に表示できない，あるいは検索できないといった問題を内在している.

表 11　電子出版の工程別特徴と問題点

区分	作り手側		利用者側
工程	執筆・編集	情報加工・蓄積	情報公開（出版）
特徴	知の創造活動	文字の性質上、漢字の出現頻度数に関係無く、膨大な字形が存在する（ロングテール）	端末や閲覧するアプリケーションによって符号化文字集合の対応が異なり、内字／外字の状況が変わる
問題点	外字・異体字指示が直接行えない場合があり、ゲラでのやりとり（赤字指示）が無くならない	外字・異体字判定やデータ化方式がバラバラで、互換性を保てないリスクが高く、対応コストも高い	外字・異体字を正確に表示できない（または検索できない）場合がある

6.2 書籍のデジタル化に伴う外字・異体字問題の解決策

6.2.1 共通識別アーキテクチャ

これまでに述べた書籍のデジタル化に伴う外字・異体字問題を解決するための前提として, デジタル化される書籍で使われる各々の文字を, 様々な利用環境に依存せず, すべての利用者が共通の認識で識別できるアーキテクチャが必要となる.

この実現のため, 一文字ごとに統一された識別番号（仮に背番号とする）を設定し, かつ出版物で利活用されている主な文字集合における符号位置との対応付けを示すマトリクス（背番号テーブル）の構築を提案する．図 2 に背番号テーブルの概念図を示す.

背番号テーブルを導入することで，ある文字が各々の文字集合において内字（文字集合内に含まれる文字)なのか外字なのかその判別が容易に行えるようになる. また, 同一視される異体字のハンドリング等が共通の認識で行えるようになる.

Vol.2011-DD-81 No.6
2011/7/22

字形判定情報				出版物で用いられる文字集合（案）						
背番号	字形サンプル（画像 128×128）			AJ1-6	UCS	IVS	凸版	大日本	文字鏡	大漢和
	字形1	字形2	字形3							
P000001	亜	亜	亜							
P000002	唖	唖	唖							
P000003	娃	娃	娃							

図 2　背番号テーブルの概念

6.3　背番号テーブルに基づく外字・異体字利用環境

図 3 に背番号テーブルに基づく外字・異体字利用環境案を示す．利用環境は，大きくは社会インフラとしての「字形共通基盤」部分と「ビジネス領域」に分けられる．字形共通基盤は，誰もが利用できるように整備される必要があり，ビジネス領域はマーケットニーズに応じたビジネスとして対応することを想定している．以下に想定環境の概要を示す．

(1)背番号テーブル
　字形一文字ごとに統一された識別番号を設定し，かつ出版物で利活用される主な文字集合の符号位置が対応づけられたテーブル
(2)字形サンプル
　背番号テーブルに登録される文字の形を示し，利用者の視覚的な共通認識を図ることを目的としたデータ
(3)文字属性テーブル
　背番号テーブルに登録された各々の文字に対する関連情報（読み、部首、画数、異体字関係等のメタデータ）を登録したテーブル
(4)(5)入力ツール，検索エンジン
　利用環境の端末において，利用者が外字・異体字を考慮すること無く文字の入力および検索ができるように支援するツール
(6)背番号と各文字集合の対応テーブル
　背番号テーブルの背番号とビジネス用途(例えば出版等)に用いる各文字集合との対応及び専用外字等の対応を関連づけたテーブル
(7)商用フォント
　利用者が出版物を利用するときに表示されるフォント
(8)外字作成ツール
　大規模な商用フォントでは吸収しきれない文字を表示するためのツール

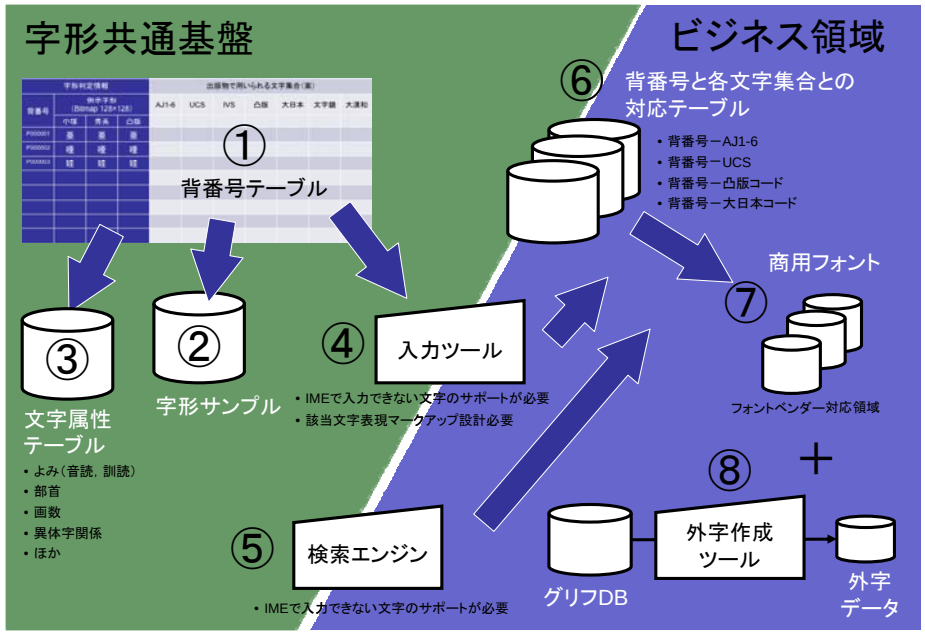

図 3　背番号テーブルに基づく外字・異体字利用環境案

7.　まとめと今後の課題

本稿では，経済産業省で行われている出版物の利活用促進のための外字・異体字利用環境整備プロジェクトについて，プロジェクトの方向性を検討する上で基礎的な資料として用いられた凸版印刷株式会社の漢字出現頻度調査の概要を示すとともに，プロジェクトの概要とその解決手法を示した．

Vol.2011-DD-81 No.6
2011/7/22

漢字出現頻度数調査では，外字・異体字の特性を考慮し，出現頻度数の低い漢字に着目し，約 99.6%は国際規格と整合性のある符号化方式で表現可能であり，それ以外の約 0.4%がユニークな名前を持つ図形で表現する必要があることを示した．

また，プロジェクトにおいては，頻度数調査の結果を踏まえ，また印刷会社・デジタルコンテンツ配信事業者・大規模プロジェクトへのヒアリングを通して作り手側・利用者側という 2 側面を考慮する利用環境アーキテクチャを示した．

現在，同プロジェクトは「平成 22 年度書籍等デジタル化推進事業」の一環として，利用環境の実証実験に着手したフェーズにある．今後は実証実験を通じてその有用性を検証する予定である．

謝辞 本研究は経済産業省平成 22 年度書籍等デジタル化推進事業の受託を受け，凸版印刷が推進しているプロジェクトである．本プロジェクトにご協力頂いている皆様に，謹んで感謝の意を表する．

参考文献

1) 漢字出現頻度数調査（3），文化庁文化部国語課(2007).
2) 知的財産戦略本部コンテンツ強化専門調査会（第 4 回）資料 2-2, http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/contents_kyouka/2011/dai4/siryou2_2.pdf
3) 文字鏡研究会, http://www.mojikyo.org/
4) インデックスフォント研究会, http://www.indexfont.com/
5) GT プロジェクト（T フォントプロジェクト）, http://charcenter.t-engine.org/tfont/index.html
6) CHISE プロジェクト, http://kanji.zinbun.kyoto-u.ac.jp/projects/chise/
7) 漢字データベース, http://kanji-database.sourceforge.net/
8) グリフウィキ, http://glyphwiki.org/wiki/GlyphWiki:%E3%83%A1%E3%82%A4%E3%83%B3%E3%83%9A%E3%83%BC%E3%82%B8
9) 文字情報基盤構築事業, http://ossipedia.ipa.go.jp/article/9/
10) 高田智和，小林正行，間淵洋子，大島一，西部みちる，山口昌也: JIS X 0213:2004 運用の検証，大規模汎用日本語データベースの構築とその活用に関する調査研究, LR-CCG-09-01，国立国語研究所(2009).