Vol.2012-CH-93 No.2
2012/1/27

網羅性を志向しない異体漢字対応テーブル

高田智和[†]　盛思超[†]　山田太造[††]

いわゆる「異体字」の概念を漢字の派生関係と通用関係とに整理した上で、人間文化研究機構研究資源共有化統合検索システムでの運用を想定し、検索のための必要最低限の「異体字」群を収録した異体漢字対応テーブルの作成事例を報告する。

A Non-exhaustive Table of Kanji Variants

Tomokazu Takada[†] Sheng Sichao[†] and Taizo Yamada[††]

This paper discusses the concept of kanji variants in terms of relationships based on formal derivation, and customary usage, and reports on the production of an optimized table listing kanji variants for use with the Resource Sharing System for the Humanities, with excludes sets of kanji variants, particularly customary usage kanji variants, that can lead to unexpected search results.

1. はじめに

日本語表記は文字レベル、語レベルともにゆれが大きく、日本語情報処理にとって大きな課題となっている。文字のレベル、特に漢字は種類もバリエーションも豊富で、異体字処理がいつもつきまとう。

かつて、田嶋一夫（1984）は異体関係にある漢字の集積と記述の必要性と、日本語情報処理での利用を述べた。その後、各処において機械可読の異体字表が作られ、入力や検索、データ結合などの便をはかっている。

検索を念頭に置いた場合、検索漏れを防ぐことを第一目的とすれば、異体字表の規模は大きい方がよいであろう。しかし一方で、ひっかかり過ぎもユーザーにとっては負担となる。検索目的に応じて、適度な規模の異体字表、あるいは、必要最低限のものに限った異体字表などがあってもよいように思われる。

本稿は、人間文化研究機構研究資源共有化統合検索システム（http://humanist.nijl.ac.jp/GlobalFinder/cgi/Start.exe）での利用を想定して、必要最低限の「異体字」群を収録した異体漢字対応テーブルの作成事例を報告する。

2. 表記のゆれのレベル

国立国語研究所（1983）では、表記のゆれを次の 4 種に分類している。

(1) 文字のレベルでのゆれ
(2) 語のレベルでのゆれ
(3) 文のレベルでのゆれ
(4) 文章のレベルでのゆれ

たとえば、/コウサ/は「交差」「交叉」と書かれる。これは語のレベルでのゆれである。「差」と「叉」の対立は、一般に/コウサ/という語についてだけ起こり、/チャーシュー/は「叉焼」であって、「叉」を「差」に入れ替えて「差焼」と書くことはしない。これに対して、「交差点」「交差點」は、文字のレベルでのゆれである。「点」と「點」の対立は、「訓点／訓點」「沸点／沸點」「点画／點画」「点を打つ／點を打つ」のように、「点」あるいは「點」を使うべきところであれば、すべてにおいてどちらも使われる可能性があり、どちらかを使ったとしても、発音が変わったり、意味が損なわれた

[†] 国立国語研究所
National Institute for Japanese Language and Linguistics
[††] 人間文化研究機構
National Institute for the Humanities

りすることはない。このように、字音・字義が変わらず、形だけが変わることを字体が異なると言い、本稿で述べる「異体漢字対応テーブル」では、文字のレベルでのゆれにあたる、字体が異なる漢字のバリエーションを扱う。

文字のレベルでのゆれには、漢字字体の違いのほかに、仮名字体の違いが該当する。語のレベルでのゆれには、「リンゴ／りんご／林檎」「百／一〇〇／100」と表記に用いる文字種の違い、「行なう／行う／行」「打合せ／打ち合せ／打ち合わせ」など送り仮名の違い、「遺跡／遺蹟」「憶測／臆測」といった漢字表記の違い、「こんにちは／こんにちわ」「ロウソク／ローソク」のような仮名遣いの違いなどがある。「バイオリン／ヴァイオリン」「エルサレム／イェルサレム」などの外来語表記の違いは、広い意味での仮名遣いの違いに含まれよう。「ベッド／ベット」「スイーツ／スィーツ／スウィーツ」などは、表記のゆれというよりも、外来語導入期・定着期にあっての語形のゆれであろう。語形のゆれが反映したものも含めて、日本語表記におけるゆれの中心部分は、語のレベルでのゆれである。文レベルでのゆれには、「奄美大島で CH 研究会がある」「奄美大島で、『CH 研究会』がある。」のように、読点や括弧の使用が該当する。

3. 「異体字」の考え方

「異体字」とは、分野や研究者によって概念設定が異なり、多義性を持つ術語のようである。一般には、規範的に正統の字体をもって「正字」とし、これとは異なる字体を一括して「異体字」と総称する（杉本つとむ（1978）など)。『康熙字典』に範をとる現代日本の漢和辞典では、「髙（はしご高)」を「俗字」などと字体価値を与えて「異体字」とし、康熙字典体の「高（くち高)」を「正字」とする。このように、正体―異体の対立の中で「異体字」が用いられる。このほかに、珍しい漢字や字体を指して「異体字」と言うこともある。

本稿では、正体―異体の対立で「異体字」を捉えることはしない。「正字／異体字」の概念は、時代や地域によって変わるものである。ここ 100 年を振り返っても、かつては、中国でも日本でも、康熙字典体の「權」が「正字」であった。字体の簡略化が進められ、日本では新字体の「権」が「正字」となり、中国では簡体字の「权」が「正字」になった（図 1)。ある字体を、何らかの価値観に基づいて「正／俗」などと呼ぶことは、説明上便宜的に用いるのはよかろうが、言語学的・文字学的分析に用いるのは妥当ではなかろう。「正字／異体字」「正／俗／通」「新字体／旧字体」「簡体字／繁体字」は、それぞれの枠組みで、個々の字体に付けられたラベル（字体価値）に過ぎないのである（笹原宏之ほか（2003))。

表 1　「正字」と「異体字」

	戦前の日本	現代の日本	現代の中国
正字	權	権（新字体）	权（簡体字）
		權（旧字体）	權（繁体字）
異体字	権 权	权	権

本稿では、字音・字義を同じくして形が異なる個々の字体を、「異体」または「異体漢字」と呼ぶ。同一の字種に所属する複数の字体の、その一つ一つが「異体」である。「權」「権」「权」のそれぞれが「異体」である。

現代日本語で、ハ行の子音には、無声声門摩擦音[h]、無声硬口蓋摩擦音[ç]、無声両唇摩擦音[ɸ]の 3 種の音声が現れる。[h]は母音[a] [e] [o]の前（ハ・ヘ・ホの子音)、[ç]は母音[i]の前（ヒの子音)、[ɸ]は母音[ɯ]の前（フの子音）にそれぞれ現れる。音声学・音韻論では、[h] [ç] [ɸ]は同一音素/h/が具現化した「異音」と言う。また、社会言語学では、言語形式のバリエーションを「変異」と言う。「異体」は、「異音」「変異」になぞらえたものである。

字種と字体は階層構造となり、図 1 のようになる（便宜的に「権」を代表させて《 》に入れて字種を表わし、〔 〕に入れて異体をそれぞれ示している)。このモデルでは、「別な字」というのは字種レベルで異なるということであり、「同じ字」というのは字種レベルで同一とみなせるということである。もちろん、字種レベルで同じか否かは、時代・地域・分野などの位相によって違いがあろう（高田智和（2009))。

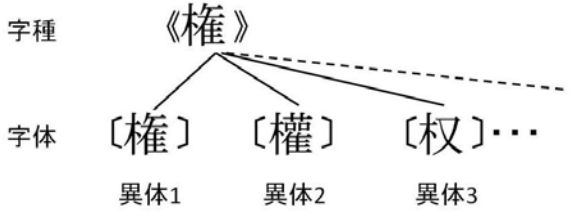

図 2　字種と字体

上述の表記のゆれのレベルに還元すると、「交差」「交叉」の違いは、「差」と「叉」が字種レベルで異なるので、語のレベルでのゆれとなり、「交差点」「交差點」の違いは、「点」と「點」が字体レベルで異なる異体同士であるから、文字のレベルでのゆれとなる。

さて、漢字の異体は、その成り立ちから、大きく 2 種類に分けることができると考えられる。派生によるものと、通用によるものである。

派生による異体は、点画の増減・省略・付加、部分字体の配置の変更（動用）などによって生じたものである。「省文」や「略体」などと呼ばれる字体は、おおよそこれに該当するであろう。

図 3 は、字種《学》を例に、派生関係を図示したものである。書体（篆書体、隷書体、楷書体など）の違いによって字体も変わる。石塚晴通（1984）は、字体の上位に書体を据え、書体が決まると字体が決まるとしている。明朝体は楷書体の一種と考えられ、〔學〕は楷書体の字体を明朝体で写し取ったものである。〔學〕〔斈〕は、肉筆で見られる字体で、〔學〕→〔斈〕の順に〔學〕から変形・派生していった字体と考えられる。現代日本の通用字体である〔学〕は、草書体の字体を楷書体化・明朝体化したものであろう。また、字種《学》には〔斉〕につくる字体がある。これも〔學〕の変形過程において生じたものとみられ、明朝体では、字体上部の「文」の 4 画目に、筆押さえの有無のデザイン差がある。こういった明朝体のデザインの違いは、字体の違いとは考えない。

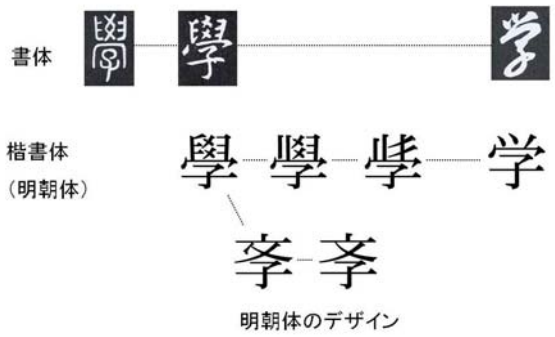

図 3　派生関係

派生関係による異体は、何か一つの発生源を想定して、書体変遷や書承過程での筆記経済によって変形し、バリエーションとして生み出されたと考えられるものであるが、異体にはもう一つ、発生源を等しくしないが、使われていくうちに同一字種に属するものと認識されるようになり、異体同士となったと考えられるものがある。通用関係の異体である。

たとえば、「万」と「萬」について、『大漢和辞典』を検すると、次のように記述されている。

> 万　西域では萬の数を表わすに卍を用いる。万の字は其の卍の変形である。
> 萬　象形。さそりの形に象る。又、仮借して、数字の万の意に用いる。

本来は別字であったものが、音が共通するため、「萬」が「万」の意味も共有するようになり、異体同士として同一字種と認識されるようになった過程が、辞書記述からうかがえる。

通用関係の形成には、音を共有する 2 字（あるいはそれ以上）の間において、一方が他方に代用されるようになり、やがて意味も共有するようになるに至って、異体関係にある（同一字種である）と一般に認識されるようになる、「万／萬」のような場合や、音・義を共有し類義の関係にある 2 字において、代用が起こり、やがて異体関係にあると認識されるようになる場合などが想定される。

派生関係であれ、通用関係であれ、異体関係が辞書記述などによってそれらが確認できたからといって、異体関係が汎時代的、汎地域的、汎分野的なものであると考えるのは避けるべきである。「著」と「着」は派生関係にあるが、現代日本でこの 2 字を同一字種と考えるのは一般的ではないだろう。「机」と「機」は、現代中国では簡体字・繁体字で、通用関係によるとみられるが、現代日本では別字種である。異体を考える上で、時代・地域・分野などの枠組みの設定は必須であり、時代・地域・分野などを限定しなくては、本来的には異体の記述は行なえないのである。

4. 統合検索システム向け異体漢字対応テーブルの作成

4.1 作成方針と手順

人間文化研究機構研究資源共有化統合検索システムでは、機構内の 6 機関（国立歴史民俗博物館、国文学研究資料館、国立国語研究所、国際日本文化研究センター、総合地球環境学研究所、国立民族学博物館）が公開しているデータベースを中心に、横断検索を提供している。異体漢字を同一視して検索する、いわゆる異体字検索はまだ実装されておらず、今回は、統合検索システムで利用する異体漢字対応テーブルの作

成を試みた。

検索システムに参加しているデータベースは、2011 年 12 月現在で 124 を数え、今後も増加が見込まれる。利用者の便をはかるために、異体字検索機能を追加することが課題となっている。また、中国関係のデータベースも含まれているため、簡体字と繁体字を同一視して検索したいとの要望もある。

個々の参加データベースが対象とする時代・地域・分野は多岐にわたるが、実際の横断検索対象となるメタデータは、目録データが中心で全文データは少なく、利用者も日本国内からが多いようであるので、対象とする異体漢字も、現代日本の標準的なものを扱い、これに中国で標準的な簡体字・繁体字を加えて、異体漢字対応テーブルを編集する方針を立てた。

具体的には、日中の公的な文字表から、それぞれの異体を抽出し、現代日本における一般的な字種の認識あわせて、テーブルを編集する。日本の文字表には、常用漢字表（2010 年）、戸籍法施行規則別表第二（2010 年）、JIS X 0213:2004 を用い、中国の文字表には簡化字総表（1964 年）を用いる。機械可読のテーブルとし、記述にあたっては Unicode を用い、字体処理には統合規則を適応させる。

以下、各文字表からの異体の抽出と、テーブルの編集について述べる。

4.2 常用漢字表

常用漢字表は、現代日本における漢字使用の目安を示した文字表で、公用文、出版、放送、教育など社会の各方面で大きな影響を持っている。2,136 字種を収録し、363 字種については、新字体に添えていわゆる旧字体を示している。新字体・旧字体の対応が 1：1 となるのは 362 字種で、1 字種《弁》のみが、新字体 1〔弁〕・旧字体 3〔辨瓣辯〕の 1 対多対応となっている。363 字種 728 字体のすべてが、Unicode で表現可能であるが、x を添えた 63 字体は、互換漢字の符号位置に対応するので、適切に表現されない場合がある。

亜亞　悪惡　圧壓　囲圍　医醫　為爲　壱壹　逸逸x　隠隱　栄榮　営營　衛衞
駅驛　謁謁x　円圓　塩鹽　縁緣　艶艷　応應　欧歐　殴毆　桜櫻　奥奧　横横
温溫　穏穩　仮假　価價　禍禍x　画畫　会會　悔悔x　海海x　絵繪　壊壞　懐懷
慨慨x　概概x　拡擴　殻殼　覚覺　学學　岳嶽　楽樂　喝喝x　渇渴　褐褐x
缶罐　巻卷　陥陷　勧勸　寛寬　漢漢x　関關　歓歡　観觀　気氣　祈祈x　既既x
帰歸　亀龜　器器x　偽僞　戯戲　犠犧　旧舊　拠據　挙擧　虚虛　峡峽　挟挾
狭狹　郷鄕　響響x　暁曉　勤勤x　謹謹x　区區　駆驅　勲勳　薫薰　径徑　茎莖
恵惠　掲揭　渓溪　経經　蛍螢　軽輕　継繼　鶏鷄　芸藝　撃擊　欠缺　研硏
県縣　倹儉　剣劍　険險　圏圈　検檢　献獻　権權　顕顯　験驗　厳嚴　広廣
効效　恒恆　黄黃　鉱鑛　号號　国國　黒黑　穀穀x　砕碎　済濟　斎齋　剤劑
殺殺x　雑雜　参參　桟棧　蚕蠶　惨慘　賛贊　残殘　糸絲　祉祉x　視視x　歯齒
児兒　辞辭　湿濕　実實　写寫　社社x　者者x　煮煮x　釈釋　寿壽　収收
臭臭x　従從　渋澁　獣獸　縦縱　祝祝x　粛肅　処處　暑暑x　署署x　緒緒
諸諸x　叙敍　将將　祥祥x　称稱　渉涉　焼燒　証證　奨奬　条條　状狀　乗乘
浄淨　剰剩　畳疊　縄繩　壌壤　嬢孃　譲讓　醸釀　触觸　嘱囑　神神x　真眞
寝寢　慎愼　尽盡　図圖　粋粹　酔醉　穂穗　随隨　髄髓　枢樞　数數　瀬瀨
声聲　斉齊　静靜　窃竊　摂攝　節節x　専專　浅淺　戦戰　践踐　銭錢　潜潛
繊纖　禅禪　祖祖x　双雙　壮壯　争爭　荘莊　捜搜　挿插　巣巢　曽曾　痩瘦
装裝　僧僧x　層層x　総總　騒騷　増增　憎憎x　蔵藏　贈贈x　臓臟　即卽
属屬　続續　堕墮　対對　体體　帯帶　滞滯　台臺　滝瀧　択擇　沢澤　担擔
単單　胆膽　嘆嘆x　団團　断斷　弾彈　遅遲　痴癡　虫蟲　昼晝　鋳鑄　著著x
庁廳　徴徵　聴聽　懲懲x　勅敕　鎮鎭　塚塚x　逓遞　鉄鐵　点點　転轉　伝傳
都都x　灯燈　当當　党黨　盗盜　稲稻　闘鬪　徳德　独獨　読讀　突突x　届屆
難難x　弐貳　悩惱　脳腦　覇霸　拝拜　廃廢　売賣　梅梅x　麦麥　発發　髪髮
抜拔　繁繁x　晩晚　蛮蠻　卑卑x　秘祕　碑碑x　浜濱　賓賓x　頻頻x　敏敏x
瓶甁　侮侮x　福福x　払拂　仏佛　併倂　並竝　塀塀x　餅餠　辺邊　変變
弁辨瓣辯　勉勉x　歩步　宝寶　豊豐　褒褒　墨墨x　翻飜　毎每　万萬　満滿
免免x　麺麵　黙默　弥彌　訳譯　薬藥　与與　予豫　余餘　誉譽　揺搖　様樣
謡謠　来來　頼賴　乱亂　覧覽　欄欄x　竜龍　隆隆x　虜虜x　両兩　猟獵　緑綠
涙淚　塁壘　類類x　礼禮　励勵　戻戾　霊靈　齢齡　暦曆　歴歷　恋戀　練練x
錬鍊　炉爐　労勞　郎郎　朗朗x　廊廊x　楼樓　録錄　湾灣

4.3 戸籍法施行規則別表第二

戸籍法施行規則別表第二は、子の名付けに使うことができる漢字を示した表で、いわゆる人名用漢字の表である。

異体を持つ字種のうち、常用漢字表字種と重ならないものは、以下の 17 字種であり、すべて異体が 2 種ずつある（17 字種 34 字体）。これらは Unicode で表現可能であるが、x を添えた 5 字体は、互換漢字の符号位置に対応するので、適切に表現されない場合がある。

亘亙　凜凛　尭堯　巌巖　晃晄　桧檜　槙槇　渚渚x　猪猪x　琢琢x　祢禰
祐祐x　禄祿　禎禎x　穣穰　萌萌　遥遙

また、常用漢字表字種ではあるが、常用漢字表の新字体・旧字体のいずれでもない異体を示しているものは、次の 8 字種である（8 字種 8 字体、下線）。いずれも Unicode で表現可能である。

園薗　駆駈　島嶋　杯盃　富冨　峰峯　野埜　涼凉

4.4 JIS X 0213:2004

JIS X 0213:2004 は、日本国内の情報交換用文字符号を定めた工業規格である。各区点位置について、参照区点位置を記述しており、これが異体関係を示しているとも考えられる。しかし、それらが標準的なものであるか否かは検討を要するので、1983 年及び 2004 年の規格改正で問題となった字種（区点位置入替字、83JIS 互換字、UCS 互換字）を抽出の対象とする（高田智和ほか（2009））。常用漢字字種、人名用漢字字種と重ならないものは、次の 58 字種である。x を添えた 1 字体は、Unicode の CJK 統合漢字拡張 B 領域に符号位置があるため、適切に表現されない場合がある。

〔区点位置入替字〕21 字種 42 字体
侭儘　壷壺　撹攪　梼檮　涛濤　潅灌　砿礦　砺礪　竃竈　篭籠　薮藪　蕊蘂
蝿蠅　蛎蠣　諌諫　賎賤　迩邇　靭靱　頚頸　鯵鰺　鴬鶯

〔83JIS 互換字〕28 字種 56 字体
侠俠　唖啞　噛嚙　嚢嚢　填塡　屡屢　掻掻　掴摑　攅攅　溌潑　涜瀆　焔焰
祷禱　箪簞　繍繡　莱萊　蒋蔣　蝉蟬　蝋蠟　躯軀　醤醤　醗醱　頬頰　顛顚
騨騨　鴎鷗　鹸鹼　麹麴

〔UCS 互換字〕9 字種 18 字体
倶俱　剥剝　叱叱 x　呑吞　嘘噓　妍妍　屏屛　并幷　繋繫

4.5 日本の文字表からの異体抽出のまとめ

日本の文字表から抽出した異体は、全部で 886 字体であり、446 字種にまとめられる（表 2）。1 字種 2 字体となるものは 429 字種 858 字体、1 字種 3 字体となるものは 8 字種 24 字体、1 字種 4 字体となるものは 1 字種である。

表 2　日本の文字表からの異体抽出

	字種	字体
常用漢字表	363	728
戸籍法施行規則別表第二	17	42
JIS X 0213	58	116
計	446	886

4.6 簡化字総表

中国の政府による文字改革は第 2 次世界大戦後に進められ、1964 年に簡化字総表にまとめられている（遠藤紹徳（1985））。簡化字総表は第 1 表、第 2 表、第 3 表から成り、第 1 表は特定の字種について簡化字を使うもの 352 例、第 2 表は部分字体として用いたときに簡化字化のパターンとなるもの 132 例（別に部分字体だけを示すもの 14 例）、第 3 表は第 2 表のパターンを応用できるもの 1,903 例が収録されている。

しかし、各表内、あるいは複数表にわたって重複掲出があるため、これを整理すると（初出を採用）、第 1 表は 352 字種、第 2 表は 132 字種、第 3 表は 1,752 字種となる。

以下、表ごとに異体抽出状況を述べる。なお、この作業にあたっては、中国文字改革委員会編『簡化字総表（第二版）』（文字改革出版社 1964 年 9 月第二版影印）を用いた。

4.7 簡化字総表第 1 表

第 1 表には 352 字種を収録するが、内訳は、簡体字・繁体字の対応が、簡体字 1 繁体字 1 となるもの 339 字種、簡体字 1 繁体字 2 となるものが 10 字種〔干乾幹、获獲穫、纤縴纖、苏蘇囌、坛壇罈、团團糰、系係繫、脏贓髒、只隻祇、钟鐘鍾〕、簡体字 1 繁体字 3〔复復複覆、蒙矇濛懞、台臺檯颱〕となるものが 3 字種である。

今回の異体漢字対応テーブルは、日本語母語話者を主体とする統合検索システムでの利用を想定しているので、現代日本において明らかに別字種となるものは、抽出しないこととする。

まず、簡体字 1 繁体字 1 対応の 339 字種のうち、簡体字・繁体字のそれぞれが、常用漢字表、戸籍法施行規則別表第二で別字種扱いとなっているものは抽出対象外とする。19 字種あり、現代中国では通用関係にあるとみられるが、現代日本では通用関係にないものと判断する。

丑醜　合閤　后後　据據　卷捲　里裏　怜憐　了瞭　面麵　舍捨　象像　叶葉
郁鬱　折摺　征徵　致緻　制製　筑築　庄荘

簡体字 1 繁体字 1 対応の 339 字種のうち、簡体字が常用漢字表、戸籍法施行規則別表第二に収録されていて、繁体字が収録されていないものは抽出対象外とする。32 字種あり、これらも現代中国では通用関係にあるとみられるが、現代日本では通用関係にないものと判断する。

板閞　表錶　別彆　卜蔔　才纔　出齣　淀澱　迭叠　冬鼕　斗鬥　谷穀　胡鬍
回迴　家傢　借藉　克剋　困睏　蔑衊　朴樸　千韆　秋鞦　曲麯　晒曬　沈瀋

松鬆　向嚮　旋鏇　踊踴　御禦　症癥　朱硃　准準

次に、簡体字 1 繁体字多対応の 13 字種については、常用漢字表、戸籍法施行規則別表第二での収録状況を考慮し、かつ、1 対 1 対応になるように調整して、次の 9 字種 18 字体だけを抽出することにした。

复復　获獲　纤纎　苏蘇　台臺　坛壇　团團　脏臟　钟鐘

したがって、第 1 表から抽出したものは、以下の 297 字種 594 字体である。検索時にこれらを同一視して差し支えないかは、なお熟考を要する。

碍礙　肮骯　袄襖　坝壩　办辦　帮幫　宝寶　报報　币幣　毙斃　标標　补補
蚕蠶　灿燦　层層　搀攙　谗讒　馋饞　缠纏　忏懺　偿償　厂廠　彻徹　尘塵
衬襯　称稱　惩懲　迟遲　冲衝　础礎　处處　触觸　辞辭　聪聰　丛叢　担擔
胆膽　导導　灯燈　邓鄧　敌敵　枭梟　递遞　点點　电電　独獨　吨噸　夺奪
堕墮　儿兒　矾礬　范範　飞飛　坟墳　奋奮　粪糞　凤鳳　肤膚　妇婦　复復
盖蓋　赶趕　个個　巩鞏　沟溝　构構　购購　顾顧　刮颳　关關　观觀　柜櫃
汉漢　号號　轰轟　壶壺　沪滬　护護　划劃　怀懷　坏壞　欢歡　环環　还還
伙夥　获獲　击擊　鸡鷄　积積　极極　际際　继繼　价價　艰艱　歼殲　茧繭
拣揀　硷鹼　舰艦　姜薑　浆漿　桨槳　奖獎　讲講　酱醬　胶膠　阶階　疖癤
洁潔　仅僅　惊驚　竞競　旧舊　剧劇　惧懼　开開　垦墾　恳懇　夸誇　块塊
亏虧　腊臘　蜡蠟　兰蘭　拦攔　栏欄　烂爛　累纍　垒壘　类類　礼禮　隶隸
帘簾　联聯　炼煉　练練　粮糧　疗療　辽遼　猎獵　临臨　邻鄰　岭嶺　庐廬
芦蘆　炉爐　陆陸　驴驢　乱亂　么麼　霉黴　梦夢　庙廟　灭滅　亩畝　恼惱
脑腦　拟擬　酿釀　疟瘧　盘盤　辟闢　苹蘋　凭憑　扑撲　仆僕　启啓　签籤
牵牽　纤纎　窍竅　窃竊　寝寢　庆慶　琼瓊　权權　劝勸　确確　让讓　扰擾
热熱　认認　洒灑　伞傘　丧喪　扫掃　涩澀　伤傷　声聲　胜勝　湿濕　实實
适適　势勢　兽獸　书書　术術　树樹　帅帥　苏蘇　虽雖　随隨　台臺　态態
坛壇　叹嘆　誉譽　体體　荣耀　铁鐵　听聽　厅廳　头頭　图圖　涂塗　团團
椭橢　洼窪　袜襪　网網　卫衛　稳穩　务務　雾霧　牺犧　习習　戏戲　虾蝦
吓嚇　咸鹹　显顯　宪憲　县縣　响響　协協　胁脅　亵褻　幷羣　兴興　须鬚
悬懸　选選　压壓　盐鹽　阳陽　养養　痒癢　样様　钥鑰　药藥　爷爺　医醫
亿億　忆憶　应應　痈癰　拥擁　佣傭　忧憂　优優　邮郵　余餘　吁籲　誉譽
渊淵　园園　远遠　愿願　跃躍　运運　酝醞　杂雜　赃贓　脏臟　凿鑿　枣棗
灶竈　斋齋　毡氈　战戰　赵趙　这這　证證　钟鐘　肿腫　种種　众衆　昼晝
烛燭　桩樁　妆妝　装裝　壮壯　状狀　浊濁　总總　钻鑽

4.8 簡化字総表第 2 表

第 2 表は 132 字種を収めるが、第 1 表と違い、派生関係のものが大部分であるため、第 1 表ほど抽出の困難はない。しかし、簡体字 1 繁体字 2 となるものが 7 字種〔当當噹、发發髮、汇匯彙、尽盡儘、历歷曆、卤鹵滷、罗羅囉〕、簡体字と繁体字との対応が現代日本では別字種となるものが 1 字種〔云雲〕あるため、調整を要する。

第 2 表から抽出したものは、次の 131 字種 262 字体である。

爱愛　罢罷　备備　贝貝　笔筆　毕畢　边邊　宾賓　参參　仓倉　产産　长長
尝嘗　车車　齿齒　虫蟲　刍芻　从從　窜竄　达達　带帶　单單　当當　党黨
东東　动動　断斷　对對　队隊　尔爾　发發　丰豐　风風　冈岡　广廣　归歸
龟龜　国國　过過　华華　画畫　汇匯　会會　几幾　夹夾　戋戔　监監　见見
荐薦　将將　节節　尽盡　进進　举舉　壳殼　来來　乐樂　离離　历歷　丽麗
两兩　灵靈　刘劉　龙龍　娄婁　卢盧　虏虜　卤鹵　录録　虑慮　仑侖　罗羅
马馬　买買　卖賣　麦麥　门門　黾黽　难難　鸟鳥　聂聶　宁寧　农農　齐齊
岂豈　气氣　迁遷　佥僉　乔喬　亲親　穷窮　区區　啬嗇　杀殺　审審　圣聖
师師　时時　寿壽　属屬　双雙　肃肅　岁歲　孙孫　条條　万萬　为為　韦韋
乌烏　无無　献獻　乡鄉　写寫　寻尋　亚亞　严嚴　厌厭　尧堯　业業　页頁
义義　艺藝　阴陰　隐隱　犹猶　鱼魚　与與　郑鄭　执執　质質　专專

4.9 簡化字総表第 3 表

第 3 表は 1,752 字種を収録する。第 2 表と同様に、簡体字 1 繁体字 2 となる 3 字種〔摆擺襬、弥彌瀰、恶惡噁〕、簡体字と繁体字との対応が現代日本では別字種となる 1 字種〔机機〕を調整した結果、1,751 字種 3,502 字体を抽出した（用例は省略）。

また、簡化字総表から抽出した異体は 4,358 字体であり、2,179 字種にまとめられる。内訳を表 3 に示す。

表 3　簡体字総表からの異体抽出

	字種	字体
第 1 表	297	594
第 2 表	131	262
第 3 表	1,751	3,502
計	2,179	4,358

4.10 2 群の結合

日中の文字表から、それぞれ取り出した 2 群の異体を結合する。日本の文字表から抽出した異体に、中国の文字表から抽出した異体を加えていく。具体的な手順は、次のとおりである。

(1) 日本の文字表から抽出した字種と、中国の文字表から抽出した字種を比較し、日本側に字種がなければ追加する。
(2) 字種が共通する場合、異体も一致するかどうかを検討し、日本側にない異体ならば追加する。

(2)について、「権權杈」を例にすると、日本の〔権權〕と中国の〔杈權〕は字種として共通する。字体レベルでは、〔權〕は日中双方にあるが、〔杈〕は日本にないので、〔権權〕に〔杈〕を加えて、結合表は〔権權杈〕となる。

結合の結果、2,369 字種 4,891 字体の結合表が得られた（表 4）。1 字種 2 字体が 2,205 字種、1 字種 3 字体が 159 字種、1 字種 4 字体が 16 字種である。また、今回作成した表では、1 行に並ぶ異体が、同一字種であることを示すにとどめ、親文字（代表字体）を立てるようなことは行なっていない。したがって、データベースの正規化などの用途で用いるには不向きである。

表 4　異体漢字対応テーブル（一部）

整理番号	異体 1	Unicode1	異体 2	Unicode2	異体 3	Unicode3	異体 4	Unicode4
064C7	択	629E	擇	64C7	抒	62E9		
064CA	撃	6483	擊	64CA				
064CB	挡	6321	擋	64CB				
064D3	扰	39DF	擾	64D3				
064D4	担	62C5	擔	64D4				
064DA	拠	62E0	據	64DA				
064E0	挤	6324	擠	64E0				
064E7	挙	6319	擧	64E7	举	4E3E	舉	8209
064EC	拟	62DF	擬	64EC				
064EF	挨	6448	擯	64EF				
064F0	拧	62E7	擰	64F0				
064F1	搁	6401	擱	64F1				
064F2	挪	63B7	擲	64F2				
064F4	拡	62E1	擴	64F4	扩	6269		
064F7	擷	64B7	擷	64F7				
064F9	摊	644A	攤	64F9				
064FA	摆	6446	擺	64FA				
064FB	撳	64DE	攇	64FB				
064FE	扰	6270	擾	64FE				
06504	摅	6445	攄	6504				
06506	撵	64B5	攆	6506				
0650F	拢	62E2	攏	650F				
06514	拦	62E6	攔	6514				

5. おわりに

本稿では、人間文化研究機構研究資源共有化統合検索システムでの利用を想定して、中国簡体字をも同一視させることを念頭に、日中の公的な文字表から標準的な異体を抽出することで、必要最低限の検索用異体集合の抽出を試みた。簡化字総表第 1 表からの抽出では、日本で別字種となる可能性があるものを排除しきれておらず、また、日本の文字表からの抽出では、「渕」や「邉」などの著名な異体が収録できていないなど、文字選定について課題を残している。今後は、実運用を通して改善を行なっていきたい。

付記　本研究は、平成 22 年度～平成 25 年度日本学術振興会科学研究費補助金基盤研究（B）「漢字字体変容の原理―敦煌文献から現代日本戸籍漢字まで―」（研究代表者：高田智和、課題番号：22320087）による成果の一部である。

参考文献

1) 田嶋一夫：漢字シソーラスの構想と課題，日本語学，3 巻 3 号，明治書院 (1984)
2) 国立国語研究所：国立国語研究所報告 75　現代表記のゆれ，秀英出版　(1983).
3) 杉本つとむ：杉本つとむ日本語講座 1　異体字とは何か，桜楓社 (1978).
4) 笹原宏之, 横山詔一, エリク=ロング : 現代日本の異体字―漢字環境学序説―, 三省堂 (2003).
5) 高田智和：常用漢字と「行政用文字」，新常用漢字表の文字論，勉誠出版 pp.55-64 (2009).
6) 石塚晴通：圖書寮本日本書紀研究編，汲古書院 (1984)
7) 高田智和，北村雅則，間淵洋子，小林正行，西部みちる，山口昌也：JIS X 0213:2004 運用の検証，国立国語研究所 (2009).
8) 遠藤紹徳：早わかり中国簡体字，国書刊行会 (1985).