

平成 30 年 6 月 27 日現在

機関番号：62618

研究種目：基盤研究(B) (一般)

研究期間：2014～2017

課題番号：26284066

研究課題名(和文) 字体記述のデジタル化に基づく文字規範史の定位

研究課題名(英文) Hanzi Normative Glyphs and Digitalised Method

研究代表者

高田 智和 (TAKADA, TOMOKAZU)

大学共同利用機関法人人間文化研究機構国立国語研究所・言語変化研究領域・准教授

研究者番号：90415612

交付決定額(研究期間全体)：(直接経費) 12,200,000円

研究成果の概要(和文)：宋版出現以前の文字規範を示す長安宮廷写経や開成石経について、「漢字字体規範史データベース」収録資料の一部を対象として、Unicode、IDS (Ideographic Description Sequence)、グリフウィキによるデジタル記述を行い、文字オントロジー (CHISE) に取り込み、「CHISE-IDS HNG漢字検索」として一般公開を開始した。また、IIIF (International Image Interoperability Framework) による字形用例整理に着手し、文字規範史・字体編年を捉える資料体構築における方法論の検討を行った。

研究成果の概要(英文)：We developed the Hanzi normative glyph database using CHISE Ontology and the glyph digitalised method, Unicode, IDS (Ideographic Description Sequence), glyph wiki.

研究分野：日本語学

キーワード：漢字字体 字体規範 字体編年 漢字字体史規範データベース 宮廷写経 開成石経 宋版 古活字

1. 研究開始当初の背景

研究代表者(高田)と文献学分野の分担者(石塚・豊島)らは、これまで日本・中国・韓国など東アジア漢字圏の漢字字体について、歴史的変遷と共時的異化の二面からなる資料体「漢字字体規範データベース(HNG)」(<http://joao-roiz.jp/HNG/>)を2004年度以来十年にわたって構築しており、このデータベースは2006年6月に「東洋文字文化の継承と発展に寄与する優れた業績」であるとして、第1回「立命館白川静記念東洋文字文化賞」を受賞した。また、その成果論文集として2012年11月には『漢字字体史研究』(勉誠出版)を上梓した(中国語版『敦煌学・日本学統編』(上海辞書出版社)は印刷中)、これらの資料体・論文集により、時代と地域により文字規範が変遷することが明示され、過去も現在も「正字」は「正字」であり続けたとする汎時代的認識などは払拭されたものと信ずる。

「漢字字体規範データベース」は、東アジア漢字圏における文字規範の時代的・地域的変遷と異化の全体像を提示するとともに、特に敦煌本を始めとする唐代以前の中国古写本と、奈良・平安時代の日本古写本を通して、初唐の標準字体が日本の標準字体として移入・定着する様相を精緻に記述する基盤を提供している。しかし、唐代に続く宋版以降の元版や明版、朝鮮古活字版が、日本の鎌倉時代から近世初期の書写本・印刷本にもたらした字体規範の変容を記述するには、基礎資料がまだ不足している。また、書写本と印刷本、すなわち筆写文字と印刷文字の字体規範の違いや影響関係の記述の精緻化にあたっては、基礎資料が十分に提供されているわけではない。

さて、近年諸処の機関・研究グループで、漢字字体の通史の記述に寄与する資料体が作成され、研究利用が可能となっている。「漢字字体規範データベース」のほかに、「拓本文字データベース」(京都大学人文科学研究所、<http://coe21.zinbun.kyoto-u.ac.jp/djvuchar>)、「木簡画像データベース・木簡字典」(奈良文化財研究所、<http://www.nabunken.go.jp/Open/mokkan/mokkan2.html>)、「電子くずし字典データベース」(東京大学史料編纂所、<http://www.hi.u-tokyo.ac.jp/ships/>)など東洋学・考古学・歴史学分野の成果がある。これらは良質な一次資料に基づき、原資料の文字画像を利用した画像データベースである。これらの画像データベースを文字の史的研究で高度に活用するためには、画像による用例提示だけでなく、用例字形そのもののデジタル記述法の確立が重要である。既存の文字コード(JIS漢字やUnicode)による用例字形の再現が一般的であるが、精緻な字形記述には粒度が不十分であり、これを補完するものとして、IVS(文字コード枝番による

字形指示)、IDS(字画などの文字構成要素への分解)、グリフウィキ(大規模デジタル字形集合)などが考案されており(次頁図参照)、漢字字体の通史の記述に足るような資料体への適用・検証を試みる環境が、情報工学分野において急速に整いつつある。文字画像データベース群の連携や横断検索など高度利活用のために、異体字集約の方式を明示した上で、国際的・学際的な合意を得たデジタル記述法を確立することが必要である。

2. 研究の目的

本研究は、本文テキスト・文字表記・書物の様式等が確立した中国宋版以降の印刷文化の影響下において、日本の文字規範が如何に形成され変遷してきたのかを実証的に明らかにするために、鎌倉時代から近世初期の日本の漢字(中国印刷本の模写本及び日本印刷本での用例整理に基づく)と同時代の中国・韓国の漢字(公的性格・規範性の高い印刷本での用例整理に基づく)について、文字規範の影響関係と史的変遷を通覧する資料体を構築し、文字規範の編年を確立する。また、本研究は、文献資料に出現する文字字形のデジタル記述法を検討し、異体字集約の方式を明示した上で国際的・学際的な合意形成を行い、資料体の高度利活用をはかることを目的とする。

また、研究期間内に下記2点を実施することを目標とする。

(1) 資料体の構築と字体編年の確立

宋版以降の元版や明版、朝鮮古活字の影響による日本の鎌倉時代から近世初期の書写本・印刷本における文字規範の形成・変容を展望するための資料体を構築し、字体編年を確立する。

(2) 文字字形のデジタル記述法の検討

本研究において構築する資料体に対してIVS、IDS、グリフウィキなどのデジタル記述法を適用し、異体字集約の方式を提示する。また、古典籍の漢字を記述するためのデジタル記述法について、字体編年をふまえた上で、欧州やアジアの文字研究者、情報工学者を交えて国際的・学際的な合意を得る。

3. 研究の方法

本研究は大別して次の三要素からなり、逐次並行して進行する。

(1) 資料体の整備・構築

本研究では、日本の文字規範の形成と変容に影響を与えた宋版、元版・明版、朝鮮古活字版と、鎌倉時代～近世初期の書写本・印刷本を選定して、印刷技術出現以降の文字規範の史的変遷と転換を展望する資料体(東アジア漢字圏印刷字体データベース)を構築する。宋版、宋版模写、日本古活字版に関しては、「漢字字体規範データベース(HNG)」

(<http://joao-roiz.jp/HNG/>)が、本研究組織の前身によりすでに15資料を構築済である。

また、元版・明版、五山版、朝鮮古活字版に関しては、資料を選定し、原本撮影・画像電子化・文字画像の切り出しと電子化本文との対応付けなどを行って資料体を増補し、「漢字字体規範データベース」と連携する。

(2) 字体編年の確認

文献に出現する漢字が規範を強く意識したものの否かは、文献の性格(公的文献か私的文献か)と書誌(料紙・版式・装丁など)に大きく関わる。文献の性格を見極めるためには、書誌学的知見に基づく原本調査が必要不可欠である。本研究では、構築する資料体の収録文献(宋版、元版・明版、朝鮮古活字版と鎌倉時代～近世初期の書写本・印刷本)について原本調査に基づき、中国・韓国と日本での文字規範の異同と影響関係を明らかにし、鎌倉時代～近世初期の日本での字体編年(文字規範の史的変遷)を確立する。このような実証的手法によって「ものさし」を導き出すことで、文献の年代推定や真贋判定へ応用することも可能となる。

本研究で字体編年を確立する区分は、宋版と宋版模写、元版・明版と五山版、朝鮮古活字版と日本古活字版の三つであり、これによって、現在の「漢字字体規範データベース」において字体編年が密に記述されている奈良・平安時代から時代を下って、近世初期までの日本の字体編年を通覧することができるようになる。

(3) 文字字形のデジタル記述の検討

古典籍の文字字形がコンピュータで再現しきれないことは、人文系諸分野においては周知のことである。しかし、近年、情報工学分野から既存の文字コード(JIS漢字やUnicode)を補完するものとして、種々のデジタル記述法が考案され実装されつつある。IVS(文字コード枝番による字形指示)、IDS(字画などの文字構成要素への分解)、グリフウィキ(大規模デジタル字形集合)などのデジタル記述法を用い、「漢字字体規範データベース」収録の宋版資料と宋版模写資料の文字字形の記述を試みる。デジタル記述の検討にあたっては、字体編年の確認をふまえた上で、国内外の文字研究者・文献研究者・情報学研究者を交えた研究セッションを設定し、異体字集約の方式を策定して、古典籍の文字字形のデジタル記述法について国際的・学際的合意を得るよう努める。これによって、文字画像データベース群の連携や横断検索など高度利活用への道が開かれる。

4. 研究成果

(1) 字体編年確立のための原本調査

字体編年確立のため、主として中世印刷本の原本書誌調査を行った。調査先は、秋田県

立図書館、東洋文庫、京都大学附属図書館、京都大学人文科学研究所、漢喃研究所、大英図書館である。

秋田県立図書館では日本五山版『韻府群玉』の原本書誌調査及び撮影を行った。また、京都大学附属図書館では中世関東印刷である普濟寺版の原本調査を行い、一部について用例整理を行った。

(2) 資料体の整備・構築

本研究期間内において、IIIF(International Image Interoperability Framework)による画像公開・画像利用が日本国内において急速に広まったため、本研究における資料体の整備・構築でもIIIFの利用に着手した。永崎研宣(連携研究者)が所属する人文情報学研究所の技術支援の下、フランス国立図書館が所蔵する長安宮廷写経『P2195 妙法蓮華経巻六』『P2413 大樓炭経巻三』、朝鮮活字本『白雲和尚抄録祖直指心體要節巻下』を対象に、同館が公開するIIIF画像による用例整理を行った。

また、「漢字字体規範史データベース(HNG)」の前身である字体資料(情報カードのスキャン画像及びメタデータ)の整理を行い、データセットリポジトリの作成を行った。

(3) 文字字形のデジタル記述法の検討

宋版出現以前の書写による文字規範(初唐標準字体)を示す長安宮廷写経について、「漢字字体規範史データベース」収録資料の一部を対象として、Unicode、IDS(Ideographic Description Sequence)、グリフウィキによるデジタル記述を行い、守岡知彦(研究分担者)が運営する文字オントロジー(CHISE)に取り込み、「CHISE-IDS HNG 漢字検索」として一般公開を開始した(<http://www.chise.org/hng-ids-find>)。

この知見を日本語学会2016年度春季大会で発表(高田智和・守岡知彦「CHISEによる漢字字体のデジタル記述 漢字字体規範史データベースを例として」)し、日本語学会大会発表賞を受賞した。

また、Unicode、IDS(Ideographic Description Sequence)、グリフウィキによるデジタル記述は、初唐標準字体以後の中国文字規範を示す開成石経(京都大学人文科学研究所蔵拓本)にも適用し、文字オントロジー(CHISE)への取り込みを行った。

(4) シンポジウム開催と図書刊行

国立国語研究所共同研究プロジェクト(基幹型)「文字環境のモデル化と社会言語科学への応用」(代表:国立国語研究所・横山詔一)科学研究費補助金基盤研究A「歴史的文字に関する経験知の共有資源化と多元的分析のための人文・情報学融合研究」(代表:奈良文化財研究所・馬場基)人間文化研究機構連携研究「表記情報と書誌形態情報を加えた日本語歴史コーパスの精緻化」(代表:

国立国語研究所・高田智和)との連携により、シンポジウム「「字体と漢字情報」 HNG 公開 10 周年記念」(2015 年 11 月 21 日 - 22 日)を開催し、日本語学・文献学だけでなく、歴史学・考古学・仏教学・心理学・情報学・日本語教育学など、分野横断的に字体と漢字情報について議論を行った。

このシンポジウムの内容をとりまとめ、論文集『漢字字体史研究 2 字体と漢字情報』としてまとめ、2016 年 11 月に刊行した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 12 件)

守岡知彦、項書き換え系を用いた漢字字体の包摂規準の形式化の試み、情報処理学会論文誌、査読有、59(2)、2018、pp.332-340

豊島正之、キリシタン日本語文典の典拠問題と電子化テキスト、訓点語と訓点資料、査読有、140、2018、pp.122-108

石塚晴通、高田智和、漢字字体と文献の性格との関係 「漢字字体規範史データベース(石塚漢字字体資料)」の文献選定、石塚晴通監修、高田智和・馬場基・横山詔一編『漢字字体史研究 2 字体と漢字情報』、査読無 2016、pp.349-359

守岡知彦、CHISE による HNG データ収録の試み、石塚晴通監修、高田智和・馬場基・横山詔一編『漢字字体史研究 2 字体と漢字情報』、査読無、2016、pp.185-203

高田智和、漢文訓読と日本語、専修大学図書館編『日本語の風景 文字はどのように書かれてきたのか』、査読無、2015、pp.163-194

高田智和、古典籍の翻刻と文字コード、国史編纂委員会『東亜細亞\$#63884;史資料共有를위協力方策模索』、査読無、2015

Tomohiko MORIOKA、Multiple-policy Character Annotation based on CHISE、Journal of the Japanese Association for Digital Humanities、査読有、1(1)、2015、pp.86-106
http://doi.org/10.17928/jjadh.1.1_86

豊島正之、キリシタン版の Vulgata 聖書引用に就て、上智大学国文学科紀要、査読無、32、2015、pp.1-14

豊島正之、キリシタン文献の表記の揺れと統一、上智大学国文学論集、査読無、48、2015、pp.54-55

高田智和、小助川貞次、古典籍原本画像と翻字テキストの対象ビューアの作成と教育利用事例、国立国語研究所論集、査読有、8、2014、pp.129-140
<http://doi.org/10.15084/00000546>

高田智和、ヲコト点の座標表現、国立歴史民俗博物館研究報告、査読有、192、2014、pp.171-181

石塚晴通、従紙材看敦煌文献的特征服、敦煌研究、査読有、敦煌研究院成立七十周年記念専号、2014、pp.118-122

[学会発表](計 27 件)

守岡知彦、CHISE における漢字部品整理の現状と課題、東洋学へのコンピュータ利用第 29 回研究セミナー、2018

石塚晴通、書体と字体 真草千字文を例として、東洋学へのコンピュータ利用第 29 回研究セミナー、2018

高田智和、福山雅深、堤智昭、小助川貞次、資料画像公開・利用の国際化と高度化の取り組み、日本語学会 2017 年度春季大会、2017

守岡知彦、CHISE の RDF 化の試み、第 114 回人文科学とコンピュータ研究会、2017

守岡知彦、前近代の漢字字形に対する字体の包摂モデルの適用に関する諸問題、第 115 回人文科学とコンピュータ研究会、2017

守岡知彦、漢字部品記述における複数ドメイン導入の試み、人文科学とコンピュータシンポジウム 2017、2017

守岡知彦、漢字構造記述再考、東洋学へのコンピュータ利用第 28 回研究セミナー、2017

福山雅深、堤智昭、高田智和、コーパス検索結果からの画像参照方式の改良、「通時コーパス」シンポジウム 2017、2017

高田智和、守岡知彦、CHISE による漢字字体のデジタル記述 漢字字体規範史データベースを例として、日本語学会 2016 年度春季大会、2016

高田智和、行政用文字の符号標準化の試み、第 61 回国際東方学者会議、2016

守岡知彦、CHISE-wiki における HNG カード画像利用の試み、第 111 回人文科学とコンピュータ研究会、2016

高田智和、東アジアの文字と文字コード シンポジウム「漢字文化圏の 100 年+」

2016

守岡知彦、長安宮廷写経の漢字字体と包摂規準 HNG と CHISE の統合を通じて、東洋学へのコンピュータ利用第 27 回研究セミナー、2016

高田智和、日下部漢字表への林大氏の書入れについて、韓国日本語文化学会 2015 年度春季国際学術大会、2015

堤智昭、田島孝治、高田智和、点図情報入力支援ツールによるヲコト点図の電子化、人文科学とコンピュータシンポジウム 2015、2015

守岡知彦、漢字の包摂粒度の符号化に関する諸問題について、第 107 回人文科学とコンピュータ研究会、2015

守岡知彦、CHISE による HNG データ収録の試み、シンポジウム「字体と漢字情報」 HNG 公開 10 周年記念、2015

守岡知彦、多粒度漢字構造モデルに基づく字形整理の試み、人文科学とコンピュータシンポジウム 2015、2015

石塚晴通、HNG 公開 10 周年にあたり、シンポジウム「字体と漢字情報」 HNG 公開 10 周年記念、2015

守岡知彦、CHISE における漢字字体・字形粒度の整理規準について、東洋学へのコンピュータ利用第 26 回研究セミナー、2015

②④高田智和、日本の常用漢字の歴史と展望、韓国日本語文化学会 2014 年度春季国際学術大会、2014

②④高田智和、漢字字体の知識と選好 台湾日本語学習者の場合、第 4 届台日亜州未来論壇「東亜文化伝播与交流 文学、思想、語言」、2014

②④高田智和、漢字字体規範史データベース、東アジア漢語データベース・シンポジウム、2014

②④高田智和、古典籍の翻刻と文字コード、東アジア史料研究編纂協議会国際学術会議、2014

②⑤Tomokazu TAKADA、Language Issues in Japanese Academia、World Script Symposia 2014、2014

②⑥守岡知彦、古漢字データベースの要件に関する試論、第 103 回人文科学とコンピュータ研究会、2014

②⑦Tomohiko MORIOKA、Digitization of a Catalogue of Oracle Bones、Japanese Association for Digital Humanities Conference 2014、2014

〔図書〕(計 1 件)

石塚晴通(監修)、高田智和・馬場基・横山詔二(編)、勉誠出版、漢字字体史研究 2 字体と漢字情報、2016、432

〔その他〕

ホームページ等

CHISE-IDS HNG 漢字検索

<http://www.chise.org/hng-ids-find>

シンポジウム「字体と漢字情報」 HNG 公開 10 周年記念

<https://www.ninjal.ac.jp/event/specialists/project-meeting/m-2015/20151121-sympo>

6. 研究組織

(1) 研究代表者

高田 智和 (TAKADA, Tomokazu)

大学共同利用機関法人人間文化研究機構
国立国語研究所・言語変化研究領域・准教授

研究者番号：90415612

(2) 研究分担者

石塚 晴通 (ISHIZUKA, Harumichi)

北海道大学・名誉教授

研究者番号：10002289

豊島 正之 (TOYOSHIMA, Masayuki)

上智大学・文学部・教授

研究者番号：10180192

守岡 知彦 (MORIOKA, Tomohiko)

京都大学・人文科学研究所・助教

研究者番号：40324701

(3) 連携研究者

池田 証壽 (IKEDA, Shoju)

北海道大学・文学研究科・教授

研究者番号：20176093

伊藤 智ゆき (ITO, Chiyuki)

東京外国語大学・アジア・アフリカ言語文化研究所・准教授

研究者番号：20361735

永崎 研宣 (NAGASAKI, Kiyonori)

一般財団法人人文情報学研究所・人文情報学部門・主任研究員

研究者番号：30343429

山田 健三 (YAMADA, Kenzo)
信州大学・人文学部・教授
研究者番号：00221656

横山 詔一 (YOKOYAMA, Shoichi)
大学共同利用機関法人人間文化研究機構
国立国語研究所・言語変化研究領域・教授
研究者番号：60182713

(4)研究協力者

イムレ ガランボス (IMURE, Galambos)
ヴォロビヨワ ガリーナ (VOROBÉVA,
Galina)
林 立萍 (LIN, Liping)