様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　６ 月 ２７ 日現在

機関番号：６２６１８
研究種目：挑戦的研究（萌芽）
研究期間：2020～2023
課題番号：２０Ｋ２０７１１

研究課題名（和文）時代差・地域差・分野差を集積した漢字字形情報通覧基盤の開発研究

研究課題名（英文）Research and development of a dataset for historical and regional kanji variants

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・研究系・教授

研究者番号：９０４１５６１２

交付決定額（研究期間全体）：（直接経費）　　4,800,000円

研究成果の概要（和文）：「漢字字体規範史データセット」（http://www.hng-data.org/）に集積した漢字約40万字を基礎データとして、漢字字形の変種（異体字）に関して、文脈に遡った使用例の分析、中国・日本の歴代の漢字字書・辞典における文字規範との対比、各研究領域で作成された異体字表との照合を行い、横断検索など人文系データの高度連携時に参照し得る、時代差・地域差・分野差を包括した漢字字形の資料体の開発研究を行った。

研究成果の学術的意義や社会的意義
本研究の意義は次の2点である。
(1) 漢字字形の時代差・地域差を、漢字文献の実用例と、漢字字書・辞典による規範の二側面から通覧し得る資料体の構築に知見を与えること
(2) 人文学諸分野間での「異体字観」の相違を明らかにし、違いに対する認識を共有した上で、分野差による字形解釈の違いを資料体に記述する方法を提案すること

研究成果の概要（英文）：Using the approximately 400,000 kanji characters accumulated in the "Hanzi Normative glyph" (http://www.hng-data.org/) as basic data, we analyzed examples of usage of variants of kanji character back in context, compared them with headwors in old Chinese and Japanese dictionaries, and compared them with variants in the tables created in various research fields. Then we have conducted research on the development of a dataset of kanji variants in time periods and regions, which can be referred to in cross-search.

研究分野：日本語学

キーワード：漢字字体　異体字　文字情報交換

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様　式　Ｃ－１９、Ｆ－１９－１（共通）

１．研究開始当初の背景

応募者は漢字資料を中心に、典籍の文字をコンピュータで表現する（電子的に翻刻する）ことを通じて、文字データベースの開発や文献学的研究を行ってきた。Unicode の普及・実装により、かつて盛んに指摘された「文字が足りない」に対しては一定の解決がもたらされたが、資料中の文字を現代のどの文字で表現するのが妥当なのか（Unicode のどの文字コードを使うのか）という問題は、資料の性質や、研究分野特有の慣習に関わっている。自らの研究において、文字データベースやデジタル翻刻をし、また、人文学諸分野（考古学、歴史学、日本文学、日本語学など）で作成されたデータを利用し、その開発に携わる過程で、漢字字形の時代差や地域差、資料を扱う研究分野での慣習の違いを認識しなければ、今後の横断的データ連携は不可能であると確信するに至った。横断検索など人文系データの高度連携時に参照し得る、時代差・地域差・分野差を包括した漢字字形の資料体が必要である。

２．研究の目的

本研究は、「漢字字体規範史データセット」（http://www.hng-data.org/）に集積した漢字約 40 万字を基礎データとして、漢字字形の変種（異体字）に関して、文脈に遡った使用例の分析、中国・日本の歴代の漢字字書・辞典における文字規範との対比、各研究領域で作成された異体字表との照合を行い、時代差・地域差・分野差を集積した漢字字形情報を通覧する基盤構築を行い、これを通して漢字字形の理論的枠組みの深化、及び学術情報交換における字形処理の高度化・精緻化に寄与することを目的とする。

３．研究の方法

本研究は、以下の 4 要素からなり、逐次並行して進行する。

①「漢字字体規範史データセット」に集積した漢字約 40 万字から、漢字字形の変種（異体字）を抽出する。

②　『干禄字書』『玉篇（広本）』『類聚名義抄（改編本）』『和玉篇』『康熙字典』など、影響力が大きい和漢の歴代の漢字字書・辞典における文字規範（「正」「俗」「通」など）の記述を抽出する。

③ ①と②とを対比し、漢字文献の実用例から帰納した漢字字形の異体体系と、漢字字書・辞典による規範性の強い異体体系を、通史的・地域的に並行させた資料体に集積する。

④ ③と人文学諸分野（考古学、歴史学、日本文学、日本語学など）の異体字表とを照合し、何を同じ文字／違う文字と見なしてきたのかを明らかにし、史的文字に関する分野別包摂粒度を明示する。

①「漢字字体規範史データセット」から漢字字形の変種（異体字）の抽出

「漢字字体規範史データセット」（http://www.hng-data.org/）は、「漢字字体規範史データベース」のデータセット（字形画像用例とメタデータ）を、オープンデータとして提供しているものである。中国は南北朝時代から宋時代、日本は奈良時代から江戸時代までの典籍 60 点から字形 40 万字を採集しているが、異体体系の抽出はまだ行われていない。〔研究代表者高田、研究分担者守岡〕

② 和漢の歴代の漢字字書・辞典から文字規範記述の抽出

漢字字書・辞典における「正」「俗」「通」などの文字規範は、固定されたものではなく、時代と地域（国）において緩やかに変遷するものである。『玉篇（広本）』『類聚名義抄（改編本）』については、「平安時代漢字字書総合データベース」（https://github.com/shikeda/HDIC）など公開データを利用し、不足のものは本研究課題で新たに作成することで、影響力が大きい中国・日本の歴代の漢字字書・辞典における文字規範の変遷を記述する。

③ 漢字文献の実用例と辞書規範との異体体系の対比

個別の文字に着目した漢字文献の実用例と辞書規範との字体認識・異体体系の対比は、文字研究においてこれまで行われてきたが、俯瞰的・網羅的な取り組みはまだ行われていない。

④ 人文学の各領域の異体字表との照合

何を同じ文字／違う文字と見なすのか、「異体字観」の相違は字形解釈（翻刻）の違いとなって現出する。人文学諸分野における「異体字観」の相違の明らかにすることで、翻刻テキストデータの利用や、史的文字の各種字形データベース（「拓本文字データベース」「木簡庫」「電子くずし字字典データベース」「日本古典籍くずし字データセット」など）の効率的横断検索の実現に知見を与える。

４．研究成果

【2020 年度】

① 奈良文化財研究所が中心となって開発した「史的文字データベース連携検索システム」（https://mojiportal.nabunken.go.jp/）に、「漢字字体規範史データセット」が参加した。これにより、「漢字字体規範史データセット」が収録する典籍の漢字字体と、木簡、古文書、日本近世印刷本の漢字字体とを比較検討する基盤が整った。

② 日本中世印刷経と宋版印刷経について漢字字体を検討するためのデータセット作成に着手した。

③ 日本語学、文献学、歴史学、考古学、仏教学、情報学など多分野の研究者が参加するシンポジウム「字体資料共有の現在と未来」（2021 年 3 月 20 日、オンライン開催、https://www2.ninjal.ac.jp/past-events/2009_2021/event/specialists/project-meeting/m-2020/20210320/index.html）を開催し、学術情報交換における字形処理の高度化・精緻化、及びデータ共有と保存に関する議論を行った。

【2021 年度】

① 日本中近世印刷本について漢字字体を検討するためのデータセット作成に着手した。

② 人間文化研究機構の次期統合検索システム（nihuBridge）に対して、異体字同値検索向けに「異体漢字対応テーブル」を提供した。また、「異体漢字対応テーブル」の改修に着手した。

【2022 年度】

① 「漢字字体規範史データセット」の原カードを調査し、原カードそのものの保存方法を検討した。

② 「異体漢字対応テーブル」を精査し、「単漢字異体字データ対応テーブル」として人間文化研究機構の機構本部内蓄積データに提供・公開した（https://bridge.nihu.jp/accumulateddata_detail/18045149）。

③ 「略字・俗字」を中心とする異体字に関して Web 調査により使用意識調査（1 都 3 県、男女 12 世代、計 600 名）を実施した。

④ 人間文化研究機構広領域連携型基幹研究「異分野融合による総合書物学の拡張的研究」、科研費基盤研究 (A)「平安時代漢字字書総合データベースの機能高度化と類聚名義抄注釈の作成」（19H00526）、一般財団法人人文情報学研究所との共催により、研究集会 「古辞書データ共有と拡張」（2023 年 1 月 21 日、オンライン開催、https://www.ninjal.ac.jp/events_jp/20230121a/）を開催し、関連諸分野の研究者と字形処理に関して議論を行った。

【2023 年度】

① 「漢字字体規範史データセット」の原カードを国立国語研究所研究資料室に納め、長期保存を行った。

② 京都大学人文科学研究所附属人文情報学創新センター、国立国語研究所共同利用推進センター、人間文化研究機構広領域連携型基幹研究「異分野融合による総合書物学の拡張的研究」等との共催により研究集会「東洋学へのコンピュータ利用 第 37 回研究セミナー」（2024 年 1 月 19 日、国立国語研究所、http://kanji.zinbun.kyoto-u.ac.jp/seminars/oricom/2024-1.html）を開催し、関連諸分野の研究者と字形処理に関して議論を行った。。

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　1件／うち国際共著　1件／うちオープンアクセス　1件）

１．著者名 守岡知彦	４．巻 2022-CH-129
２．論文標題 CHISEにおけるHDIC統合の試み	５．発行年 2022年
３．雑誌名 情報処理学会研究報告	６．最初と最後の頁 1-6
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 藤本灯、韓一、高田智和	４．巻 21
２．論文標題 古辞書の構造化記述の試み―『和名類聚抄』を例に―	５．発行年 2021年
３．雑誌名 国立国語研究所論集	６．最初と最後の頁 85-94
掲載論文のDOI（デジタルオブジェクト識別子） 10.15084/00003438	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕　計10件（うち招待講演　2件／うち国際学会　1件）

１．発表者名 守岡知彦
２．発表標題 低品質文字画像を用いた高精細画像における字形画像の自動再切り出しの試み
３．学会等名 東洋学へのコンピュータ利用第35回研究セミナー
４．発表年 2022年

１．発表者名 守岡知彦
２．発表標題 CHISEにおけるHDICサポートの現状と課題
３．学会等名 研究集会「古辞書データ共有と拡張」
４．発表年 2023年

１．発表者名 高田智和
２．発表標題 漢字の字体・異体字
３．学会等名 日本がん登録協議会第30回学術集会（招待講演）
４．発表年 2021年

１．発表者名 守岡知彦
２．発表標題 説文小篆に対する漢字構造記述の試み
３．学会等名 東洋学へのコンピュータ利用第34回研究セミナー
４．発表年 2021年

１．発表者名 高田智和
２．発表標題 板碑の文字
３．学会等名 国立国語研究所オープンハウス2021
４．発表年 2021年

１．発表者名 Morioka Tomohiko
２．発表標題 Machine-readable description of Chinese Characters— Overview and background of CHISE —
３．学会等名 KUDH International Conference（国際学会）
４．発表年 2021年

１．発表者名 高田智和
２．発表標題 学術研究のための多言語・多文字への対応と多漢字・変体仮名
３．学会等名 日本学術会議公開シンポジウム「総合知創出に向けた人文・社会科学のデジタル研究基盤構築の現在」（招待講演）
４．発表年 2022年

１．発表者名 守岡知彦
２．発表標題 圏論に基づく漢字構造記述のモデル化の試み
３．学会等名 第128回人文科学とコンピュータ研究会
４．発表年 2022年

１．発表者名 守岡知彦
２．発表標題 漢字字体規範史データセットにおける版管理
３．学会等名 シンポジウム「字体資料共有の現在と未来」
４．発表年 2021年

１．発表者名 高田智和
２．発表標題 普済寺版の漢字字体
３．学会等名 シンポジウム「字体資料共有の現在と未来」
４．発表年 2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

漢字字体規範史データセット https://www.hng-data.org/

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	守岡　知彦 (Morioka Tomohiko) (40324701)	京都大学・人文科学研究所・助教 (14301)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関