

令和 3 年 6 月 29 日現在

機関番号：62618

研究種目：基盤研究(B) (一般)

研究期間：2018～2020

課題番号：18H00674

研究課題名(和文) 訓点資料訓読文コーパスの構築と古代日本語史研究の革新

研究課題名(英文) New developments in the historical Japanese language research based on the construction of kunten materials' corpora

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) 13,300,000円

研究成果の概要(和文)：訓点資料の精読に基づいて、訓点資料用のコーパスアノテーションを設計した。西大寺本『金光明最勝王経』を対象に、設計したコーパスアノテーションを実装・検証し、訓点資料解読文コーパスを試作した。試作したコーパスを利用して、訓点の計量分析を行った。また、形態論情報付与のため、訓点資料の語彙・語法の調査・分析を行った。

研究成果の学術的意義や社会的意義

日本語史研究は、大別して和文系資料と漢文系資料と双方を用いることで展開してきた。和文系資料は『日本語歴史コーパス』の構築・公開により資料体の整備が進んでいるが、もう一方の漢文系資料は資料体の整備が不十分である。本研究によって漢文訓点資料のコーパス構築が前進し、コーパスベースの訓点資料を用いた日本語史研究を展開させる基盤が整えられた。

研究成果の概要(英文)：Based on the careful reading of the kunten material in the early heian period, we designed the corpus annotations for the kunten materials. We implemented the designed annotations for "Golden Light Sutra (Jingguangming zuishengwang jing)" in in early middle Japanese, and constructed a prototype corpus of kunten material. Then, using the prototype corpus, we performed a quantitative analysis of the kunten. In addition, in order to provide morphological information for prototype corpus, we studied the vocabulary and syntactic conventions of the kunten materials in the early heian period.

研究分野：日本語学

キーワード：訓点資料 漢文訓読 訓読文 コーパス

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、Z - 19 (共通)

1. 研究開始当初の背景

日本語史研究では、表記(片仮名の成立、仮名遣)、音韻(音便、漢字音)、語彙・語法(漢文訓読語彙)などの史の変遷を捉えるために、漢文訓点資料が活用されてきた。特に、9世紀平安時代初期においては信頼に足る和文系資料が存在しないため、漢文訓点資料を用いた種々の研究成果が公表されてきた。大矢透『仮名遣及仮名字体沿革史料』(1909年)、春日政治『西大寺本金光明最勝王経古点の国語学的研究』(1942年)、中田祝夫『古点本の国語学的研究』(1954年)、築島裕『平安時代訓点本論考』(1986年)などである。

漢文訓点資料から、平安時代初期以降の日本語諸現象の事実と変遷とが見出され、語彙や文体の面で和文系資料とは異なる漢文訓読語彙・漢文訓読体が、歴史的に併存してきたことが明らかとなった。一方で、ヲコト点を中心とする訓点の研究は高度に発達し、あるヲコト点を使用する教学流派に対する理解なくして漢文訓点資料研究は成り立たなくなっている。このことは、漢文訓点資料の専門研究者でない日本語史分野の研究者が、漢文訓点資料の研究利用を敬遠する一因になっていると考えられる。

また、近年の『日本語歴史コーパス』の構築と公開は、日本語史分野の研究に便宜をもたらしている。しかし、『日本語歴史コーパス』の収録資料は和文系資料が主体であり、歴史的に併存してきたもう一方の漢文系資料が収録されていない。漢文系資料収録に対する期待は大きいものであるが、現時点で収録が見送られているのは、漢文訓点資料に対するデジタル技術導入の基礎研究が停滞していることにも淵源を求めることができる。

漢文訓点資料のこれまでの解読成果を共有し、日本語史研究での高度利用を実現するためには訓点資料のコーパスが有効かつ不可欠である。また、訓点資料のコーパス構築は、漢文訓点資料に対するデジタル技術導入を促進するものである。

2. 研究の目的

本研究は、漢文訓点資料の解読成果である訓読文をコーパスとして再構築し、表記・音韻・語彙・語法などの史の変遷を捉える基盤を整えた上で、訓点の定量的分析や、漢文訓読語彙と和文語彙との精緻な比較対照分析を行い、日本語学分野の史的研究を格段に発展させることを目的とする。

日本語史研究は、大別して和文系資料と漢文系資料と双方を用いることで展開してきた。和文系資料は『日本語歴史コーパス』の構築・公開により資料体の整備が進んでいるが、もう一方の漢文系資料は資料体の整備が不十分であり、漢文訓点資料の資料体は皆無である。本研究は『日本語歴史コーパス』の一部門として期待されているが、種々の困難のために最後まで残されてきた漢文訓点資料のコーパス構築を、現在望みうる最高の研究者グループを組織して遂行するものである。

また、本研究において構築する訓点資料のコーパスは、訓読文テキストに諸訓点(ヲコト点、仮名点、声点、句読点、返読点など)のタグを付与するものであり、ヲコト点や仮名点を指定した検索・用例抽出が容易となる。そのため、訓点の定量的分析や、漢文訓読語彙と和文語彙との精緻な比較対照分析の進展が期待できる。特に、前者の訓点の定量的分析には、例えば漢文訓点資料での語種比率(和語・漢語の使用比率)などが含まれる。これまで訓点の計量が困難であったため、漢文訓点資料を対象としたこの種の基礎計量もほとんど存在しない。本研究において構築する訓点資料のコーパスによって、漢文訓点資料を対象とした計量研究・コーパスベースの研究のみならず、漢文訓点資料を加えた日本語史分野の計量研究・コーパスベースの研究が現出する。

3. 研究の方法

本研究は、(1)原資料(影印及び原本)の諸訓点(ヲコト点、仮名点、声点、句読点、返読点など)を精査した上で、訓読文テキストに対する諸訓点のマークアップ方法の検討(訓点用タグセットの検討)、(2)訓読文コーパスの作成、(3)デジタル版訓点総索引のプロトタイプ作成、(4)訓点の定量的分析及び漢文訓読語彙と和文語彙との比較対照分析の4要素から成り、逐次並行して進行する。

(1) 原資料の諸訓点の精査及び訓点用タグセットの検討

「西大寺本金光明最勝王経」(830年頃の訓点)を対象に、原資料(影印及び原本)の諸訓点の精査を通して、訓点用タグセットの検討し確定する。

(2) 訓読文コーパスの作成

漢文訓点資料研究において定評のある春日政治博士の訓読文(『西大寺本金光明最勝王経古点の国語学的研究』(1942年)に収録)を対象に、訓読文コーパスの設計を行い試作する。

(3) デジタル版訓点総索引のプロトタイプ作成の検討

従来の紙媒体の訓点総索引を精査し、デジタル版訓点総索引の項目選定を行い、訓読文コーパスとの連携機能を前提とし、デジタル版訓点総索引の設計を行う。

(4) 訓点の定量的分析及び漢文訓読語彙と和文語彙との比較対照分析

訓読文コーパスやデジタル版訓点総索引を利用した訓点の定量的分析及び漢文訓読語彙と和文語彙との比較対照分析の方法を検討する。

4. 研究成果

(1) 原資料の諸訓点の精査及び訓点用タグセットの検討

『西大寺本金光明最勝王経』(830年頃の訓点)を中心に、原資料(影印及び原本)の諸訓点(ヲコト点、仮名点、声点、句読点、返読点など)と、春日政治博士の訓読文(『西大寺本金光明最勝王経古点の国語学的研究』(1942年)に収録)の精査を行い、訓点用タグセットを考案した。従来の『日本語歴史コーパス』のアノテーションになく、訓点用のアノテーションとして新たに考案したものには以下のものがある。

要素名	内容
kanji	漢字の所在位置、漢字に記された訓点を属性として表示する。
wokototen	訓読文中の当該語が原漢文ではヲコト点で記入されていることを表す。
kanaten	訓読文中の当該語が原漢文では仮名点で記入されていることを表す。
hodoku	原漢文への施点はないが、訓読時に補われた本文であることを表す。
mute	訓読文で不読字として処理した文字を表す。

また、原資料の諸訓点の精査を通して、春日政治博士の訓読文の一部を補訂しうる知見を得た。この知見に基づき、『西大寺本金光明最勝王経』巻一の新訓読文を作成した(後日、国立国語研究所サイトから公開を予定している)。

(2) 訓読文コーパスの作成

漢文訓点資料研究において定評のある春日政治博士の訓読文(『西大寺本金光明最勝王経古点の国語学的研究』(1942年)に収録)及び(1)の新訓読文を対象に、(1)の訓点用タグセットを実装・検証し、構造化テキスト(XML形式)を試作した(新訓読文を対象とした構造化テキストは、後日、国立国語研究所サイトから公開を予定している)。また、構造化テキスト(XML形式)の検用可視化ビューアの試作も行った。

(3) デジタル版訓点総索引のプロトタイプ作成の検討

従来の紙媒体の訓点総索引から代表的なものを選び、それらの項目や掲出語の言語単位を精査した。『日本語歴史コーパス』で採用されている言語単位(短単位と長単位)との一致・不一致を確認し、訓読文コーパスに対する形態論情報付与の課題を整理した。また、デジタル版訓点総索引のプロトタイプ作成に関する課題整理を行った。

(4) 訓点の定量的分析及び漢文訓読語彙と和文語彙との比較対照分析

(2)で試作した訓読文コーパスを用いて、ヲコト点・仮名点の計量分析を行った。「西大寺本金光明最勝王経」巻一の漢文本文6,427字では、ヲコト点が付された漢字60.0%、仮名点が付された漢字12.0%であった。また、ヲコト点または仮名点が付された漢字49.9%、どちらも付されていない漢字51.1%であった。このほか、1漢字1ヲコト点、1漢字2ヲコト点...などの計量や、1漢字複数ヲコト点のパターン計量、ヲコト点の使用度数など、種々の基礎計量を行った。ヲコト点の共起関係や加点位置の相補関係は、語法・構文解釈において重要であり、形態論情報付与に対する知見を得た。

(5) 成果公表活動

訓点語学会や日本語学会などにおいて成果公表を行った。また、京都大学人文科学研究所附属東アジア人文情報学研究センター・国立国語研究所研究情報発信センター等との共催により「東洋学へのコンピュータ利用第31回研究セミナー」(2019年7月26日、於国立国語研究所)、人文情報学研究センター等との共催により「シンポジウム・字体資料共有の現在と未来」(2021年3月20日、於国立国語研究所)を開催し、漢字資料のコーパス・データベース開発に携わる研究者と分野横断的に情報交換・議論を行った。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 柳原恵津子	4. 巻 19
2. 論文標題 平安初期訓点資料における不読字の再検討 コーパス・電子化テキストを用いた訓点語研究の試みとして	5. 発行年 2020年
3. 雑誌名 国立国語研究所論集	6. 最初と最後の頁 187-207
掲載論文のDOI（デジタルオブジェクト識別子） 10.15084/00002835	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 -

1. 著者名 田島孝治、堤智昭、高田智和	4. 巻 61巻2号
2. 論文標題 訓点資料の書き下し文自動生成を目的としたヲコト点を中心とする訓点の計量分析	5. 発行年 2020年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 162-170
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 有
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 高田智和・福山雅深・堤智昭・小助川貞次	4. 巻 15
2. 論文標題 資料画像公開・利用の国際化と高度化の取り組み 「日本語史研究資料 [国立国語研究所蔵] の事例	5. 発行年 2018年
3. 雑誌名 国立国語研究所論集	6. 最初と最後の頁 163-176
掲載論文のDOI（デジタルオブジェクト識別子） 10.15084/00001601	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 -

〔学会発表〕 計8件（うち招待講演 1件/うち国際学会 0件）

1. 発表者名 柳原恵津子、高田智和
2. 発表標題 訓点資料に使用されたヲコト点・仮名点の計量研究 西大寺本『金光明最勝王経』平安初期点を用いて
3. 学会等名 日本語学会2020年度秋季大会
4. 発表年 2020年

1. 発表者名 柳原恵津子、高田智和
2. 発表標題 訓点資料コーパスの設計と意義 西大寺本『金光明最勝王經』平安初期点を用いた実践例
3. 学会等名 第123回訓点語学会研究発表会
4. 発表年 2020年

1. 発表者名 柳原恵津子
2. 発表標題 『金光明最勝王經』平安初期点の形態素解析用本文作成 その方法と問題点
3. 学会等名 東洋学へのコンピュータ利用第31回研究セミナー
4. 発表年 2019年

1. 発表者名 守岡知彦、劉冠偉、高田智和
2. 発表標題 漢字字体規範史データセットと単字検索
3. 学会等名 日本語学会2019年度秋季大会
4. 発表年 2019年

1. 発表者名 高田智和
2. 発表標題 中世多摩の文字づかい 板碑と經典文字からわかること
3. 学会等名 多摩郷土誌フェア（招待講演）
4. 発表年 2020年

1. 発表者名 高田智和
2. 発表標題 『石塚漢字字体資料』と『漢字字体規範史データベース』
3. 学会等名 シンポジウム「文字情報データベースの保存と継承」
4. 発表年 2018年

1. 発表者名 堤智昭・高田智和・田島孝治・小助川貞次
2. 発表標題 訓点資料電子化の取り組み ヲコト点データベースの試作を例に
3. 学会等名 日本語学会2018年度秋季大会
4. 発表年 2018年

1. 発表者名 柳原恵津子
2. 発表標題 訓点資料の形態素解析用本文作成とその意義 「西大寺本金光明最勝王経」平安初期点を用いて
3. 学会等名 「通時コーパス」シンポジウム2019
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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	小木曾 智信 (OGISO Toshinobu) (20337489)	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 (62618)	

6. 研究組織（つづき）

	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	佐伯 俊源 (SAEKI Shungen) (30309067)	種智院大学・人文学部・教授（移行） (34308)	
研究分担者	高山 倫明 (TAKAYAMA Michiaki) (90179565)	九州大学・人文科学研究院・教授 (17102)	
研究分担者	田中 草大 (TANAKA Sota) (20778758)	京都大学・文学研究科・講師 (14301)	
研究分担者	月本 雅幸 (TSUKIMOTO Masayuki) (60143137)	東京大学・大学院人文社会系研究科（文学部）・教授 (12601)	

	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
研究協力者	柳原 恵津子 (YANAGIHARA Etsuko) (50401162)	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・非常勤研究員 (62618)	

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