

令和 2 年 7 月 10 日現在

機関番号：62618

研究種目：挑戦的研究（萌芽）

研究期間：2017～2019

課題番号：17K18506

研究課題名（和文）漢文訓点資料の国際文書構造記述による共有化と書き下し文自動生成のための基礎研究

研究課題名（英文）Computerized Method of Kunten and Quantitative Analysis of Kunten Materials for Interpreted Text Generation

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） 4,700,000円

研究成果の概要（和文）：本研究課題においては、訓点の定量的分析のための文書記述の検討、訓点入力を主目的とした移点ツールの作成、主要26点図を収録したワコト点図データベースの作成を行った。また、上記の文書記述を『尚書（古活字版第三種本）』（国立国語研究所蔵）において実践し、「尚書（古活字版第三種本）訓点情報データベース」として公開した（<https://cid.ninjal.ac.jp/kunten-syousyo3/>）。ワコト点図データベースの公開も行った（<https://cid.ninjal.ac.jp/wokototendb>）。

研究成果の学術的意義や社会的意義

本研究の意義は、デジタル技術の導入による訓点資料の解読成果の学界共有の基盤構築を推進させた点と、ワコト点を中心とする訓点研究において定量的分析を可能にした点である。

研究成果の概要（英文）：We propose a computerized method of Kunten Materials for quantitative analysis and automatic interpreted text generation. We develop a Kunten input support tool and two databases. Kunten Database (<https://cid.ninjal.ac.jp/kunten-syousyo3/>) is the Database of gloss information (Kunten information) on Shangshu (printed in the late 16th century) at the National Institute for Japanese Language and Linguistics. Wokototen-zu Database the Database of Wokototen-zu which is a reading mark chart of vernacular reading of a Chinese texts, and it can search Wokototen by the type, position, shape, meaning of the Wokototen.

研究分野：日本語学

キーワード：訓点資料 漢文訓読 ワコト点 文書構造

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

かつて漢文は社会・文化・学術を語る東アジアの共通書記言語であった。東アジアの中国周辺民族は漢文を受容し文明を育ててきた。日本も例外ではない。自言語とは異なる漢文を受容し理解するにあたって、東アジア各地で漢文訓読(自言語による理解)が行われた。最も高度に発達し長く続いたのが日本の漢文訓読である。

漢文訓読のために訓点を記入した典籍・文書を漢文訓点資料と言う。漢文訓点資料は、前近代の政治・経済・思想・宗教・文芸などあらゆる分野の典籍・文書に及び、歴史学・文学・仏教学・文化学など広く史的研究を行う上での根幹資料となっている。また、日本語史研究では、文字・音韻・語彙・語法などの史の変遷を解明するために、漢文訓点資料が活用されてきた。

しかし、本文が漢文であることに加え、訓点は複雑な書き入れであるため、漢文訓点資料研究へのデジタル技術の導入が立ち遅れている。これを解消し、人文学諸分野で漢文訓点資料の解読成果を共有するために、本研究課題を構想するに至った。

研究代表者の高田は、10年来、国内外で典籍の原本調査を行い、漢文訓点資料の漢字字体の研究と字体データベースの作成、漢文訓点資料の解読成果(釈文)の公表を行ってきた。上述の漢文訓点資料研究の問題点を解決するため、国立国語研究所において共同研究プロジェクト「訓点資料の構造化記述」(平成21年10月~24年9月)を実施し、解読結果(釈文)の構造化記述方法を検討し、原本画像に遡った検証が可能なデジタル釈文のプロトタイプを作成を行った。本研究課題は、漢文訓点資料の解読成果を学界で共有するための基盤構築を、日本語学・歴史学・仏教学・情報工学の分野横断的研究体制のもと、発展的に行うものである。

2. 研究の目的

本研究課題は、平安・鎌倉時代の漢文訓点資料を対象として、

- (1) 漢文本文及び訓点(読み仮名、送り仮名、ヲコト点・声点・句読点・語順点などの各種記号)を文書構造の国際記述(TEI: Text Encoding Initiative)に基づく構造化記述法を考案する。
- (2) 平安・鎌倉時代を特徴付ける訓点であるヲコト点(漢字の四隅・四辺等に記入して読みを表す記号)のデータベースを作成する。
- (3)(1)(2)のデータに基づいて、ヲコト点の種類と系統を定量的観点から考察する。
- (4)(1)(2)のデータに基づいて、漢文訓点資料の書き下し文の自動生成方法を検討する。

これらにより、漢文訓点資料の解読成果を学界で共有するための基盤構築を目的とする。また本研究の意義は次の2点である。

- (1) 漢文訓点資料の解読成果の共有
- (2) ヲコト点を中心とする訓点の定量的分析

(1)について、漢文訓点資料は日本語学・歴史学・文学・仏教学・文化学などにおいて、広く史的研究を行う上での根幹資料である。しかし、原本の書写面・印刷面を、影印本や複製本として図書で刊行する、あるいは、原本の本文・訓点の解読成果である書き下し文等を図書で刊行するというように、漢文訓点資料研究の成果は紙媒体で公表されるのが一般的である。近年のデジタルアーカイブの発達によって、紙媒体に代わって、原本の書写面・印刷面が画像データで提供される機会も増えてきたが、デジタル技術を用いて漢文訓点資料の解読成果を共有することはあまり進んでいない。漢文訓点資料の解読成果をデジタル的に記述する方法が、学界において未だ確立していないことが原因である。本研究課題において提案・実践する、文書構造の国際記述(TEI: Text Encoding Initiative)に基づく漢文訓点資料の記述法により、分野横断的かつ国際的に解読成果を共有するためのデジタル記述法が確立する。

(2)について、ヲコト点の体系的な研究は、中田祝夫『古点本の国語学的研究』(講談社、1954年)、築島裕『平安時代訓点本論考』(汲古書院、1996年)により一定の到達点に達したと見られる。しかし、その後も各研究者による調査研究によって新たな知見も加わっている。従来の文献学的観点による分析に加え、本研究で作成する点図データベースによって、定量的観点からのヲコト点の体系的な研究の展開が見込まれる。また、個別の漢文訓点資料に対して、仮名点(読み仮名や送り仮名など)やヲコト点などの基礎計量を行ったものはほとんどなく、本研究課題によって作成する構造化記述を施したデータを活用することで、初めて訓点の定量的分析を行うことが可能になる。本研究課題は、ヲコト点を始めとする訓点について体系的な研究の展開を図るものである。

3. 研究の方法

本研究課題では、漢文訓点資料のヲコト点の種類と系統を定量的観点から分析し、また、書き下し文の自動生成方法を検討するため、次の2点のデータを作成する。

- (1) 漢籍・仏典・国書などの各種訓点資料について、漢文本文及び訓点(読み仮名、送り仮名、ヲコト点・声点・句読点・語順点などの各種記号)を文書構造の国際記述(TEI: Text

Encoding Initiative) に基づく構造化記述法を考案・実践する。

- (2) 点図集(ヲコト点を集積した資料)及びこれまでに各研究者が訓点資料の原本調査によって記述した点図から、ヲコト点の形状、記入位置、読み方を抽出した点図データベースを作成する。

また、本研究課題は上記2点のデータ作成を中心に、研究期間内において、以下の5点を逐次平行して実施する。

- (1) 構造化記述法の考案・実践〔研究代表者：高田、研究分担者：田島、連携研究者：小倉、永崎〕
- (2) 点図データベースの作成〔研究代表者：高田、研究分担者：堤〕
- (3) ヲコト点の種類と系統の計量的分析〔研究代表者：高田、研究分担者：堤、連携研究者：小助川〕
- (4) 書き下し文の自動生成方法の検討〔研究代表者：高田、研究分担者：田島、堤〕
- (5) 研究成果の公表

4. 研究成果

本研究課題の成果は、大きく、(1) 訓点の定量的分析のための文書記述の考案、(2) 訓点入力を主目的とした移点ツールの作成、(3) 主要26点図を収録したヲコト点図データベースの作成である。

(1) については、訓点資料に記入されたヲコト点をはじめとする各種訓点を精査し、定量的分析を目的とした記述法の提案を行った。

高田智和・田島孝治・堤智昭「加点情報の再構成」『研究報告人文科学とコンピュータ 2018-CH-117』、pp.1-4、2018

上記の成果は「訓点資料研究の新たな展開を示すパイオニア的研究」との評価を得て、情報処理学会 2019 年度(令和元年度)山下記念研究賞を受賞した(<https://www.ipsj.or.jp/award/yamasita2019-detail.html#ch>)

一方で、文書構造の国際記述(TEI: Text Encoding Initiative)への対応は、本研究期間内においては検討にとどまり今後の課題を残した。

(2) については、訓点資料の初期解読で行われる「移点」(原資料の訓点の有無や意味を確定する作業。一般に原資料を見ながら転記する。)効率化のためのツールを作成し、公開した(<http://www.gifu-nct.ac.jp/elec/ktajima/tools.html>)。また、移転ツールを用い、『尚書(古活字版第三種本)』(国立国語研究所蔵)に対して(1)の文書記述を実践し、下記のデータベースを公開した。

尚書(古活字版第三種本)訓点情報データベース(<https://cid.ninjal.ac.jp/kunten-syousyo3/>)

移転ツールの作成と文書記述の実践、さらに記述結果による基礎計量や書き下し文自動生成性について、以下の論文を発表した。

林昌哉・田島孝治・堤智昭・高田智和・小助川貞次「訓点資料の加点情報計量のためのデータ構造 国立国語研究所蔵「尚書(古活字版)」を対象として」『人文科学とコンピュータ シンポジウム 2017 論文集』、pp.45-52、2017

田島孝治・Baptiste Jannequin・堤智昭・高田智和「IIIF Viewer と連携可能な訓点資料の加点情報データベースの試作」『人文科学とコンピュータシンポジウム 2019 論文集』、pp.109-114、2019

田島孝治・堤智昭・高田智和「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」『情報処理学会論文誌』第61巻第2号、pp.162-170、2020

(3) については、喜多院点や古紀伝点など点図集所載の主要26点図(築島裕『訓点語彙集成』第1巻、汲古書院、2007)を対象として、整数座標を用いてデータ化し、下記のデータベースを公開した。また、これまでに各研究者が作成した点図に対して、同じ整数座標によるデータ化を試行し、点図データと連携させる訓点資料の書誌情報データの設計も検討した。

ヲコト点図データベース(<https://cid.ninjal.ac.jp/wokototendb>)

ヲコト点図データベースの作成と拡張、点図集所載の主要26点図の計量分析について、以下の論文を発表した。

堤智昭・田島孝治・小助川貞次・高田智和「訓点資料の構造化記述方式と計算機を用いた基礎計量」『情報処理学会論文誌』第59巻第2号、pp.278-287、2018

堤智昭・田島孝治・高田智和・小助川貞次「ヲコト点図共有・比較プラットフォームの実装」『人文科学とコンピュータシンポジウム 2019 論文集』、pp.103-108、2019

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 0件）

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) http://doi.org/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 堤智昭・田島孝治・小助川貞次・高田智和	4. 巻 59(2)
2. 論文標題 訓点資料の構造化記述方式と計算機を用いた基礎計量	5. 発行年 2018年
3. 雑誌名 情報処理学会	6. 最初と最後の頁 278-287
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 田島孝治・堤智昭・高田智和	4. 巻 61(2)
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	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 -

〔学会発表〕 計8件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 高田智和・田島孝治・堤智昭
2. 発表標題 加点情報の再構成
3. 学会等名 第117回人文科学とコンピュータ研究会
4. 発表年 2018年

1. 発表者名 林昌哉・田島孝治・高田智和
2. 発表標題 尚書（古活字版）の訓点データの基礎計量
3. 学会等名 第118回人文科学とコンピュータ研究会
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 林昌哉・田島孝治・堤智昭・高田智和・小助川貞次
2. 発表標題 訓点資料の加點情報計量のためのデータ構造 国立国語研究所蔵「尚書（古活字版）」を対象として
3. 学会等名 人文科学とコンピュータシンポジウム2017
4. 発表年 2017年

1. 発表者名 堤智昭・土井裕絵・田島孝治・高田智和・小助川貞次
2. 発表標題 ヲコト点データベースと検索システムの試作
3. 学会等名 人文科学とコンピュータシンポジウム2017
4. 発表年 2017年

1. 発表者名 田島孝治・堤智昭・高田智和・小助川貞次
2. 発表標題 移点ツールの仮名点・語順点への拡張
3. 学会等名 第120回人文科学とコンピュータ研究会
4. 発表年 2019年

1. 発表者名 堤智昭・田島孝治・高田智和・小助川貞次
2. 発表標題 ヲコト点図共有・比較プラットフォームの実装
3. 学会等名 人文科学とコンピュータシンポジウム2019
4. 発表年 2019年

1. 発表者名 田島孝治・Baptiste Jannequin・堤智昭・高田智和
2. 発表標題 IIIF Viewerと連携可能な訓点資料の加点情報データベースの試作
3. 学会等名 人文科学とコンピュータシンポジウム2019
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>ヲコト点図データベース https://cid.ninjal.ac.jp/wokototendb 尚書（古活字版第三種本）訓点情報データベース https://cid.ninjal.ac.jp/kunten-syousyo3/ 訓点資料の移点ツール（岐阜工業高等専門学校田島研究室） http://www.gifu-nct.ac.jp/elec/ktajima/tools.html</p>
--

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	田島 孝治 (TAJIMA KOJI) (90611640)	岐阜工業高等専門学校・その他部局等・准教授 (53701)	
研究分担者	堤 智昭 (TSUTSUMI TOMOAKI) (80759035)	東京電機大学・情報環境学部・助教 (32657)	
連携研究者	小倉 慈司 (OGURA SHIGEJI) (20581101)	国立歴史民俗博物館・研究部・准教授 (62501)	
連携研究者	小助川 貞次 (KOSUKEGAWA TEIJI) (20201486)	富山大学・人文学部・教授 (13201)	
連携研究者	永崎 研宣 (NAGASAKI KIYONORI) (30343429)	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 (82683)	