

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 19 日現在

機関番号：41503

研究種目：若手研究(B)

研究期間：2012～2014

課題番号：24720213

研究課題名(和文) 中世漢語声点資料による画像付きデータベース構築の研究

研究課題名(英文) Study of the construction of the digital graphics database using shoten materials in Kamakura Period-Muromachi Period Sino-Japanese words

研究代表者

加藤 大鶴 (KATO, Daikaku)

東北文教大学短期大学部・その他部局等・准教授

研究者番号：20318728

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：本研究では字音声調の史的研究に役立てることを目的とした「中世漢語声点資料による画像付きデータベース」(以下DB)を作成した。字音声調研究においては、声点の物理的な位置情報によって声調を認定する作業が必須であるが、この手続は研究者の主観的な観察に任されることが多かった。本DBでは声点付き漢字の画像、透明テキストによる文脈付き画像、韻書等の研究上必要な情報を検索結果として表示させることで、声調の認定を検証可能な形で提示することができた。

研究成果の概要(英文)：This paper is about the production of the digital graphics database (hereafter DB), using materials of shoten (tone-marks) in Kamakura Period-Muromachi Period Sino-Japanese words, with the purpose of furthering historical research into jion tones (Japanese pronunciation of kanji based on adopted Chinese pronunciation). The vital process of acknowledging jion tones should be through objective observation of the physical location of tone marks, however, too often the actual procedure has been left to individual researchers' subjective observation. The DB contains graphics of Chinese characters with tone marks and graphics with context using layered transparent text, and research results are displayed together with other relevant information based on 9th-11th century Chinese rhyme dictionaries such as "Guang-yun". Therefore it allows us to present acknowledgement of jion tones in a verifiable manner.

研究分野：漢語アクセントの史的研究

キーワード：声点 データベース 漢語アクセント

1. 研究開始当初の背景

(1) ウェブ上の言語史研究データベース

近年、言語史研究のためのデータベース(以下 DB と称する)がウェブ上にあいついで公開されている。例えば、木簡画像が閲覧できる「木簡字典」、「漢字字体規範データベース」等はいまや史的な射程を持つ文字研究には欠かせないものとなっている。

これらのデータベースの特徴は、テキストデータと画像データを組み合わせる点にある。ウェブ上の DB には紙幅の制約がないために元データの画像も DB に含められ、その結果、検索の利便性とテキストデータの客観性・妥当性の検証可能性が保持されていることは言うまでもない。こうした DB が公開されるにつれて DB の設計方法や画像提示の方法論なども進展してきた。しかし声点資料の研究を目的とした画像付き DB はいまだウェブ上に公開されたものがない。

(2) 中世における漢語声点資料の蓄積

応募者は中世における漢語声点資料として、『平家物語』、『尾張国郡司百姓等解文』、『宝物集』などを調査し、データ化した。また、平成 21 年度～23 年度に科学研究費補助金・基盤(C) (研究代表者：上野和昭)「漢語アクセントの解明と資料の発掘」によって、中世の漢語声点資料を和文声点資料と比較することで、漢語アクセントの変遷を明らかにしつつある。これらの資料に現れる漢語声点は、漢籍・仏典類に比べて字音声調の体系的な知識や規範の影響を受けにくく、日本語の音調・アクセントの影響を研究する上では重要な性格を持つことを指摘してきた。

(3) 声点認定の客観性担保の問題

本応募者が調査してきた中世における漢語声点資料のデータを用いて画像付きの検索可能な DB 化すれば、画像と声点データを一覧の形式で出力することができる。これによって漢語アクセント研究に役立つデータが検証可能な形で公開することができる。文献に基づく漢語アクセント史の記述のためには声点の位置情報を認定し、そこから声調・アクセントを推定する基礎的手続きを踏まえる必要がある。このうち前者のプロセスは原則的に研究者の観察に任せられ、その信頼性・妥当性が検証可能な形で示されることはほとんどない。検証の対象となるのは誤読や誤写、誤入力だけではなく声点の認定そのものも含まれる。とりわけ、平声と入声の重点と軽点を認定するには客観性の担保が難しい。

2. 研究の目的

(1) 中世漢語声点資料の整理と拡充

本応募者がすでに調査を終えた『平家物語』・『尾張国郡司百姓等解文』・『宝物集』諸本のデータを本 DB 用に整理する。さらに平声・入声に軽点を区別する中世漢語声点資料を発掘しデータとして追加する。

(2) 画像つき漢語声点 DB の作成

本 DB の特徴はテキストデータと画像データを両方検索・出力できる点にある。検索可能な画像データは以下の 2 種類準備し、該当箇所のみ出力と前後の文脈付き画像出力とに対応させる。

差声された漢語部分のみ切り抜いた画像データ

透明テキストつき画像データ

後者については「透明テキストつき画像作成ツール(安岡孝一 2004)を用いれば、JPEG 形式に透明テキストを重ねあわせた検索可能な画像データを作成できる。これにより画像化された数行分の 1 ページ全体を表示しながら、該当箇所を検索し閲覧することができることになる。差声された漢語部分の出力はもちろんのこと、前後の句とのつながりも表示させることで漢語への差声を理解することができる。

以上の画像データをテキストデータとタグにて連携させ、漢語、声点、仮名語形などで検索できるようにする。

(3) 声点の認定に関わる問題の研究

中世漢語声点資料の整理と拡充、画像つき漢語声点データベースの作成のうえに、声点の認定に関わる問題についての研究を展開させる。具体的には、次の 2 点を目標とする。

画像を用いて軽点の認定に関わる字形上の特徴を抽出する

差声から記述される軽点を含む漢語の声調型と推定されるアクセント型を比較し検証する

では資料に記された声点の具体的位置がどのようにして声調表示機能を持てたかという問題の再検証となる。では記号から解釈される音調型が、日本語として実現したであろう音調型と矛盾なく説明できるか検証する。

3. 研究の方法

(1) 文献画像データの準備

A. 資料のデジタルデータ化

対象となる資料は、『尾張国郡司百姓等解文』『宝物集』『平家物語』等の諸本である。諸本所蔵者に DB への 2 次利用の許諾を得た上で、紙媒体の写真を必要に応じてスキ

ランしデジタルデータ化する。ウェブ上に画像が公開されている場合はそれを本 DB 用に加工する。

B. 漢語声点データの切り出し

デジタルデータ化された各資料から、漢語声点が付された部分を切り出し、個別画像データにした。後に DB 上でテキストデータと連携させるために、ファイル名の一覧も作成した。

C. 透明テキスト付き画像データの作成

「透明テキスト付き画像作成ツール (ttext-kanbun)」を用いて、透明テキスト付き画像データを作成する。資料に漢語声点が現れる位置情報を埋め込み、その位置情報と該当箇所のテキストデータを重ねあわせ、検索可能な画像データとする。

A.B.C.いずれの作業も本応募者が中心に行うが、A.B.および C.のうち位置情報埋め込み作業は大学院生の協力を得る。このほか、適宜資料の探索と補充を行う。

(2) DB の設計・構築・試験的公開

A. 画像データとテキストデータの整理

前年度に準備した画像データのファイル群を、テキストデータと連携可能な形式とした。テキストデータは 1 項目につき、主キー・資料略号・出現箇所・漢語・声点・推定アクセント型・推定語形を設定し、主キーと画像データを連携させた。

B. 画像付き漢語声点データベースの設計

ウェブサーバー上で動く MySQL を利用して、クエリを実行する DB を設計する。クエリは 漢語(漢字表記)、推定語形(仮名表記)、声調の検索を行い、単字による部分検索も可能なものとする。また、検索結果はテキストデータと画像データを並列させる。システム設計は研究協力者の依田平氏の助力を得ながら進め、DB 構築自体は業者に発注する。

C. DB の試験的公開

研究室内に配置した Linux ウェブサーバー上に DB を置き試験的に外部公開しながら、動作の確認と不具合のチェックを行った。

(3) 外部公開と DB の研究利用開始

平成 26 年度は最終年度であるので、本 DB を利用して声点の認定に関わる問題についての研究に着手し、テキストデータに現れる声点認定の客観性・妥当性の検証を画像データに基づいて実証的に行う。また、DB の本格的な外部公開を実施する。

4. 研究成果

(1) データベースの構築

本研究の大きな目的である DB の構築について述べる。DB では、声点の認定と韻

書での所属声調、日本語での拍数情報など字音声調研究に必要な情報をテキストで検索結果に示すとともに、漢字に差された画像を並べて閲覧できるようにした。また連音上の声調変化を研究するためには語構成や文脈を知ることも重要であり、漢語がどう切りだされているかという問題に対応するため、レイヤー化された透明テキストつき画像を PDF 形式にて作成し、漢語の画像検索結果とリンクさせた。すなわち当該漢語の情報を検索すると同時に本文上の文脈を参照することが可能となった。

(2) データと画像

本 DB で検索可能なデータの典拠は、以下に示す 3 種類 8 本の文献である。データの総数(漢字字数)は 2585 字。画像公開利用の許諾が得られているのは『尾張国郡司百姓等解文』早大本・東大本のみである。

A. 『尾張国郡司百姓等解文』

早大本：早稲田大学図書館蔵弘安 4 年(1281)写本。画像は早稲田大学図書館古典籍データベースに公開されたものによる。延べ 385 字。

東大本：東京大学史料編纂所蔵応長元年(1311)写本。画像は東京大学史料編纂所データベースによる。延べ 119 字。

真福寺本：真福寺宝生院蔵正中 2 年(1325)写本。延べ 49 字。

B. 『宝物集』調査はいずれも小泉弘 1973 の影印によった。

久遠寺本：身延山久遠寺蔵 15 世紀末頃写本。延べ 212 字。

光長寺本：光長寺蔵弘安 10 年[1288]写本。延べ 6 字。

C. 『新猿楽記』調査はいずれも前田育徳会尊経閣文庫 2010 の影印によった。

古抄本：書写年代不詳。延べ 732 字。

康永本：康永三年(1344)抄本。延べ 314 字。

弘安本：弘安三年(1280)抄本。延べ 768 字。

(3) データベースの運用

A. フリーワード検索

画面上部のフリーワード欄に、漢字・漢語、仮名音注、推定音形・推定語形、諸橋大漢和辞典番号のいずれかを入力することで、検索結果を得ることができる。検索結果には、漢字検索の場合では当該漢字と画像・資料名・出現箇所・声点(平/平軽/上/去/入/入軽)・仮名音注・漢字環境(語頭/非語頭)・漢字を発音した場合の拍数(1拍/2拍)・韻書における声調(平/上/去/入)・韻書における清濁(全清/

次清 / 清濁 / 全濁)の9項目が表示される。漢語検索の場合では当該漢語と画像・声点型・仮名音注・拍数と構造・字数・推定語形の6項目が表示される。また、得られた検索結果をクリックすることで、漢字から漢語、漢語から漢字へと相互に情報を参照することが可能である。検索に際して、完全一致検索をする場合は" "(ダブルクォーテーション)で囲む。" "がない場合は部分一致検索となる。

B. 検索結果の表示を変更する

検索結果は結果左上部分にあるアイコンをクリックすることで、画像+テキスト情報 テキスト情報のみ 画像のみ の形式に切り替えることができる。

C. 声点検索・韻書声調検索

本 DB では各文献に現れる声点(型)および広韻所属声調(型)から当該漢字・漢語を検索することができる。画面上部の声点 または 広韻声調・清濁 欄をクリックすると、平 平軽 上 去 入 入軽 * の7枚のカードが現れる。それらを 声点 または 広韻声調・清濁 枠に並べる(ドラッグ&ドロップする)ことで、漢字または最大で5字までの漢語の検索結果が得られる。

D. 拍数の選択、および単点・双点の選択について

カード上部をクリックすることで、漢字を発音した場合の拍数(1拍か2拍)を選択することができる。何も表示されていない状態は、拍数に関係なく検索する。

また 声点 欄に並べたカード下部をクリックすることで単点・双点(濁音を表す)を選択することができる。何も表示されていない状態は、単点・双点に関係なく検索する。広韻声調・清濁 欄では同様の動作で中国語音韻学で言う全清音・次清音・次濁音・全濁音を選択することができる。

E. 透明テキスト付画像による文脈の確認

検索結果の右半分画面をスクロールダウンし 関連する資料 から 資料を表示をクリックすると、前後の文脈を含む透明テキスト付き画像で該当箇所を確認することができる。画像に重ね合わせられているテキストは レイヤー文字を非表示 にして閲覧することも可能である。また PDFをダウンロード をクリックすると新規ウィンドウが立ち上がり、該当箇所を含む大きな透明テキストつき画像が表示される。なお各文献画像を冒頭部分から閲覧するにはトップ画面(画面左上 中世漢語声点資料による画像付きデータベース をクリックすることで移動)に一旦戻り、資料を閲覧する からプルダウンメニューを選択する。

F. CSV 形式による検索結果ダウンロード
検索結果画面の上部にある CSV から検索結果をダウンロードし、2次利用することができる。ダウンロードしたファイルは表計算ソフトなどで閲覧する。

(4)声点とアクセント型の整合性

まず本研究では、陰類平声字に差される声点が平声軽点である場合、その音調をF(1拍字)またはHL(2拍字)とみなした。陰類平声字を含む2字漢語のアクセント型を推定したところ、3拍漢語を例にすれば、FLL型、LLF型、LHF型を認めることができる。さらにそれをアクセントの体系変化後の資料と対照したところ、それぞれFLL-HLL、LLF-HLL、LHF-LHLに規則的に対応することが読み取れた。すなわち、アクセント型で見る限り、陰類平声字を下降調とみなすことに不自然はないことがわかった。ただし、陰類平声字であっても、実際に平声軽点が差されるか、平声点が差されるかは、もちろん資料によって異なっているし、同じ差声であっても異なるアクセント型を想定できる場合もある。例えば陰類平声+陽類平声で構成される2字漢語に 上平 の差声があった場合、古く中国語またはそれが伝承された音調を厳密に聞き分けたり発音したりする人の場合はFLLで把握されていたが、原音に対する厳密さを欠く人々の音調ではHLLと聞きなされ発音されていた。それを四声体系で記述すれば 上平 となる。あるいは原音に対する厳密さのために実音調がFLLであったとしても四声の枠組みで記述すればやはり 上平 と差声せざるを得なかった。ゆえに 上平 はFLLまたはHLLと解釈されることになる、といった具合である。

5. 主な発表論文等

[雑誌論文](計2件)

加藤大鶴、音韻史を担う漢語アクセント中低形回避・低起上昇型間の揺れ・原音声調との非対応例を中心に、論集(10)、査読無し、31-64、2015

加藤大鶴、中世和化漢文資料に現れる漢語声点の揺れ：『新猿楽記』弘安本・康永本・古抄本の比較から、論集(8)、査読無し、23-41、2012

[学会発表](計4件)

加藤大鶴、漢語アクセント史を担う語と担わない語 漢字2字2拍語と3拍語を例として、拡大アクセント史資料研究会、2014

加藤大鶴、中世漢語声点資料による画像付きデータベースの公開、日本語学会、

2014

加藤大鶴、『新猿楽記』の二字漢語における去声字のふるまい 漢音読み語彙と呉音読み語彙の異なり 、拡大アクセント史資料研究会、2013

加藤大鶴、漢語アクセントの伝統性と非伝統性 和化漢文・和漢混淆文資料と『平家正節』の比較から 、拡大アクセント史資料研究会、2011

〔その他〕

ホームページ等

「中世漢語声点資料による画像付きデータベース」

<http://gassan.t-bunkyo.ac.jp/kango-shouten/>

6．研究組織

(1)研究代表者

加藤 大鶴 (KATO, Daikaku)

東北文教大学短期大学部・准教授

研究者番号：20318728

(4)研究協力者

依田 平 (YODA, Taira)