科学研究費助成事業 研究成果報告書

令和 2 年 6 月 5 日現在

機関番号: 17102 研究種目: 若手研究 研究期間: 2018~2019

課題番号: 18K12399

研究課題名(和文)訓点資料本文データベース作成のためのシステム構築

研究課題名(英文)Construction of a system for a Kunten full-text database

研究代表者

蛭沼 芽衣 (HIRUNUMA, Mei)

九州大学・人文科学研究院・助教

研究者番号:20807177

交付決定額(研究期間全体):(直接経費) 1,300,000円

研究成果の概要(和文):本研究は、訓点資料本文データベースを作成することを目的に、そのためのシステムを構築することを目指す。そのために(1)データ項目の選定(2)入力フォームの作成をおこなった。システムにはExcelのマクロを使用した。多くのパソコンで使えるからである。入力データ項目には、訓点資料の特質や、これまでされてきた訓点資料の研究から、訓点資料の国語学的資料としての位置づけなどを考慮しながら入力データ項目を選定した。次に、入力項目に対し、それらの入力の助けとなるようなフォームを作成した。一部を自動入力されるようにすることで、訓点と漢文についての基礎的な理解があれば、ある程度データ入力ができるよう工夫した。

研究成果の学術的意義や社会的意義 訓点資料は、上代末から中古~中世前期にかけておこなわれた漢文訓読の成果である。後世の写本しかない和文 資料とくらべ、加点時の原本が残っているため、当時のことばとしての信頼性が高い。資料の点数も膨大で、内 容的にも、ヲコト点や仮名点による語彙、語法のみならず、声点から字音声調やアクセント、返り点などから統 語構造、仮名点の字体から仮名字体の変遷などをも知ることのできる、実に情報量の多い資料群である。しか し、調査や解読が難しく、現在の日本語研究に於て、充分に活用されているとはいいがたい。本研究でのシステ ムを基に、訓点資料データベースが作成することができれば、新たな研究可能性が開けると予想される。

研究成果の概要(英文): This research aims at system construction to create a Kunten Full-text database. So I chose an data items and created an input form. For the database, I use Excel, because it can be used on many computers. So I used macro function of Excel for input form. Select Items: The data items were selected from characteristics of Kunten materials and the research on the Kunten materials that have been conducted so far, taking into account their significance as Japanese language studies.

Input Form: Since there are many inputs items, I created an input form to make it easier to input. I created a part that can be entered automatically and devised it so that data can be entered relatively easily.

研究分野: 国語学

キーワード: 訓点資料 データベース

様 式 C-19、F-19-1、Z-19(共通)

1.研究開始当初の背景

訓点資料とは、漢文を訓読する際に付された「訓点(仮名点、ヲコト点)」をもつ資料群のことである。加点は上代末から始まり、中古~中世前期に盛んにおこなわれた。後世の写本でしか残存していない和文資料と違い、加点時の原本が残っているため、当時のことばとしての信頼性が高い。資料の点数も膨大で、いまだ未調査の資料も各所に眠っているものと思われる。

訓点資料は、ヲコト点や仮名点によって、 語彙や 語法が示されるのみならず、 声点から字音声調やアクセントを、 返り点などから統語構造を、 仮名点の字体から仮名字体の変遷などをも知ることができる実に情報量の多い資料群である。しかし、データ化はおろか、出版・公開されているものが少なく、利用しにくいのが現状である。そこで訓点資料に示される様々な情報を兼ね備えた訓点資料本文データベースを構築したいと考えた。

2.研究の目的

訓点資料本文データベースを構築するにあたって、本研究では、まず、訓点資料本文データベースのためのシステムづくりを目指す。

訓点資料は上述のように、情報量が多い。入力を少しでも簡便にし、データ収集の効率化をはかる必要がある。また、多くの場合、入力事項が同じことがあるので、入力のシステム化が可能になると考えた。しかも訓点資料の解読には、ある程度の知識と経験を要する。入力方法のデザインを工夫することで、少し練習をした学生でも、データが扱えるようにしたい。つまり

- (1) 訓点資料を使用した日本語研究に資することのできるようなデータセットを有するデータベースであること。
- (2) データベースの作成には、わかりやすく、システマ化されているものを使えるようにすること。

が本研究で作成する予定のデータベースの目標である。

3.研究の方法

上述のように、訓点資料は膨大な数眠っているが、解読には知識が必要となり、容易に扱える 資料ではない。そのうえに、データ入力に専門的な知識が必要となれば、データベースの作成は 困難となる。そこで、ある程度訓点資料の知識を持つ者なら、だれでもデータ収集ができるよう、 Excel のマクロ機能を使用することにした。

以上を基に、本研究では、

- (1) 入力データ項目の選定
- (2) 入力システムの作成

をおこなった。

(1)入力データ項目の選定

データベースの構築に際し、どのような情報を備えているべきかを選定する。訓点資料の特質や、これまでされてきた訓点資料の研究から、訓点資料の国語学的資料としての位置づけなどを考慮しながら入力データ項目を選定した。また、その時の入力情報のあり方も工夫した(研究成果(1)参照)。

さらに、大学院の演習授業において、院生に協力してもらい、実際にデータを作成しながら、フィードバックを返してもらった。演習の準備段階で、自分の担当範囲についてのデータを作成してもらいつつ、どの項目がどのような形式で示されるとわかりやすいか、どこが自動入力化されると便利かなど、多くの意見をもらうことで、項目の整理をしていった。

(2)入力システムの作成

入力を簡便に行えるよう、ユーザーフォームを作成し、視覚的に操作しやすく、入力内容が分かりやすいものを目指した。

訓点資料は、現在知られているだけでも数千点にのぼるが、未調査のもの、存在が知られていないものを合わせれば何万点にもなると予想される。加点量は資料にもよるが、上記の入力項目の内容によって、入力に多くの時間を要するものもあると予想される。上述のように、「ある程度訓点を解せるものであれば誰でも」を目標に、感覚的に使えるようなデザインを考える必要がある

これには、研究協力者を雇い、このシステムを使用しながら入力作業を行ってもらい、必要(不要)なシステムについての意見をだしてもらった。同時にプログラムのデバッグもしてもらい、不具合やフォームデザインについて改良を加えていった。

4. 研究成果

(1)データ項目

データの単位は漢字一字とし、それぞれの漢字について、設定した項目のデータを収集するようにした。入力データの項目は、 資料中における漢字の位置情報(No)、ヲコト点の 形、 位置、 形と位置から導き出されるヨミ、仮名点の 字体情報、 ヨミ情報、声点の 形、 位置、 単声点か複声点かの別、 声点が字音として差されているのか真言陀羅尼に差されているのかの別、 その他の符号(句読や音訓符号など)の形と その意味、またそれぞれの 墨色である。

3		N	o.		T								ヲコ	卜点		Т	1	反名点		Т		声	in the		Т	符	号		語彙		
4	No.	巻 7	「 行 ▼	字 ~	本	割 *	漢字	確認・	点番	墨.		位置	補膝・	3 %	総合ヨミ	#	ヨミ情勢	字体情報	補設	# # ·	形**	単複	声点	梵漢 *	果.	形。	意味	返点		備考	-
6635	6631	2		3	2	0	偏	偏に		1 白	1	Н		=	=																
6636	6632	2		3	3	0	袒	袒力タヌき		1 🛱	1	С		+	+	白	カタヌ	カ1タ13	1												
6637	6633	2		3	4	0	右	右		0																					
6638	6634	2		3	5	0	肩	屑を		1 🗎	1	D		7	7										白	1	懸				
6639	6635	2		3 (6	0	右	右の		1 🛱	1	В		1	1																

データ形式

ヲコト点が複数ある場合は、ひとつづつにわけ、記録するようにした。そのため、複数のヲコト点を足しあわせた「総合ヨミ」の項目も加えてある。点の記述には、形ごとに番号を振り、ヲコト点の配置にも記号を与えて対応した。仮名点の字体情報も同様に、仮名字体ごとに番号を振ったが、微妙な差異のものもあるので、これについては、今後も検討していく必要があると考えている。字体は加点年代の目安となるため、それを反映できる形での記述のしかたを考えていかなければならない。



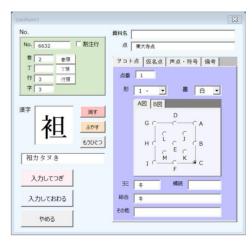
また、現在はデータ収集の便宜上、漢字一字を単位としているが、これでは不十分であることがわかった。漢語の形態素は基本的に字と対応するが、「菩薩」など、明らかに二字以上で一語となっているものもある。そのため、「薩は」だけをデータとして集めても意味がないのである。また、今後の日本語研究の資料として使えるようにするために、品詞や活用の種類などの情報を増やしていきたいと考えている。漢字一字をベースとしつつも、そこから単語を抽出し、幅広いデータにしていくシステムを作り出すことを、今後の課題としたい。

(2)入力システム

入力システムは、Excel のユーザーフォームを使用した。 ヲコト点の形と位置を指定すると、自動的にヨミが入力されるようにした。こうすることによって、位置や形が不確かな場合でも、文脈にあわせて検討していくことができる。このようにして、訓点と漢文についての基礎的な理解があれば、ある程度データが集められるようにしている。

データ項目は多いが、一字あたり、すべての項目が必要となることはほとんどないので、ヲコト点、仮名点など、点の種類ごとのフォームも作成した。

それでも、項目の多さから煩雑さはまぬがれない。また、機能として必要であっても、エクセルのシステム上できないこともある。このような場合、かえってフォームによる入力が手間となることもあることがわかってきた。入力用のインターフェースについても、まだ検討の余地がある。また、上述のように、項目を増やした場合の入力や整理の方法についても、フォームの調整をしていかなければならない。今後の課題としたい。



入力フォーム

5 . 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

_

6 . 研究組織

 ・ 101 フ しか丘が現		
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