HOKKAIDO UNIVERSITY

Title	日本漢字辞書研究の資料と方法に関する基礎的研究［論文内容及び審査の要旨］
Author(s)	劉, 冠偉
Degree Grantor	北海道大学
Degree Name	博士(文学)
Dissertation Number	甲第14572号
Issue Date	2021-03-25
Doc URL	https://hdl.handle.net/2115/81135
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	Guanwei_Liu_abstract.pdf, 論文内容の要旨

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文内容の要旨

博士の専攻分野の名称：博士（文学）　　氏名：　劉　冠　偉

学位論文題名
日本漢字辞書研究の資料と方法に関する基礎的研究

・本論文の観点と方法

日本の漢字辞書は長い歴史があり、膨大な漢字情報と和訓を収録している。国語学、訓点語と訓点資料の研究には不可欠な材料である。近年、日本の古辞書に関する発表・研究が増え、活発化する傾向が見える。これは、日本の古辞書に関するデータベース・研究ツールが拡充されてきたためである。また、コンピュータとインターネットの普及によって、研究資料としての文献画像・テキストが公開され、研究環境が大きく変化した。日本の漢字辞書の場合、“平安時代漢字字書総合データベース”や“日本語史研究用テキストデータ集”などのテキストデータが公開され、辞書に掲載されている漢字と和訓の量的な研究が可能になった。これらを利用するためには、日本の古辞書に関して、次のような問題を解決することが必要である。

① 注文の構造が複雑で、漢字の形・音・訓・漢文義注と引用出典の分析が十分ではない。
② 写本資料ごとに、符号化方針が異なり、異体字・形近異字の検索が難しい。
③ 簡単に検索・表示できるインターフェースがなく、応用のシーンが限られている。

本研究は、これらの新しい日本漢字辞書研究資料を分析して、情報学の手法を最大限にとりこみ、従来の人文学（国語学）の成果を検証し、新たな研究領域を開拓することを目指したものである。具体的には、上記の三つの問題に対して、次のように解決に向けた道筋を立てた。

① 注文項目構造の多様性に対応したタグセットとマークアップ・ツールを開発する。
② 漢字字書テキストデータから和訓と異体字関係を抽出して、対応関係を分析する。
③ ウェブインターフェースを開発して、成果を利用性高い研究資料として公開する。

ただし、本研究の提案は、実際の資料へ応用する際に、次の三つの条件を満たす必要がある。

● 国語・日本語の研究に有益なもの
● テキストデータが公開されたもの
● インターネット公開され、権利関係上問題がなく、利用しやすいもの

以上の条件を満たす資料として、平安時代の観智院本『類聚名義抄』、室町時代の夢梅本『和玉篇』、明治時代の上田万年『大字典』の三つの辞典のテキストデータを利用した。

・本研究の内容

本論文は、序論と結論を入れて、全 7 章からなる。第 2 章から第 6 章までは本論に当たる。以下、各章の内容を述べる。

第 1 章では、序論として研究の背景、目的、方法、利用資料、論文の構成を述べる。

第 2 章では、各利用資料のデータ形式を述べ、基礎統計を行った。各資料の漢字と和訓との対応関係、和訓の分布を算出したうえ、三つの辞典における共通する掲出字とそれらの和訓を抽出し、共通する掲出字―和訓の組み合わせをまとめた。

第 3 章では、部首分類体の日本古辞書の項目構造は、掲出字と注文からなるが、注文は大別して音注、意義注、字体注、和訓の 4 要素からなる。音注は字音注、意義注は漢文意義注、義注、漢文義注と呼ぶことがある。さらにこの 4 要素には多様な形式で注記が施され、それぞれの要素に数種のタイプが認められる。さらに訓点、傍訓などデータ化しにくい情報が存する。したがって、日本古辞書のデータベース化には、項目構造の多様性に対応したタグセットとマークアップ・ツールの開発が求められる。本章では、観智院本『類聚名義抄』を対象にして、タグセットとそれに対応させた日本古辞書マークアップ・ツール「tagzuke」について述べる。

第 4 章では、構造化した注文を用いて、各資料に対してさらに高度な分析を行う。前章で述べたように、観智院本『類聚名義抄』は複雑な項目構造を持っている。掲出字は字数により、単字の形式（単字形式）と 2 字以上の掲出字の形式（複字形式）に分けられる。複字形式はさらに熟語に関する項目（熟語項目）、異体字に関する項目（異体項目）、その他の項目の三種類に分けることができる。本章では、主に前章で構造化した注文データを用いて、観智院本『類聚名義抄』の項目の自動分類と分析を行った。

第 5 章では、日本漢字辞書における字体と字形に関する研究である。前半では、観智院本『類聚名義抄』を例にして、研究資料を公開する際に、漢字における符号化による制限および著作権による制限を克服するため、漢字字形共有データベース GlyphWiki による字形の再現方法を述べた。再現した字形は原本画像の代用としてデータベースのインターフェースで表示することが可能となる。さらに、GlyphWiki で字形を記述する Kage データと漢字部品構成データを比較して、原本の字形と翻刻された符号との相違を観察した。後半では、中国辞書の宋本『玉篇』のテキストデータを構造化して、異体字情報を抽出した。

第 6 章では、研究成果を公開するプラットフォーム HDIC Viewer の開発について述べる。HDIC Viewer は最初 HDIC（平安時代漢字字書総合データベース）を公開・検索するためのインターフェースとして開発された。その後、HDIC 本文の校訂・新技術の検証が増え、HDIC 以外の古辞書データベースを追加した。現在は hdic2. let.hokudai.ac.jp で公開している。現時点では、HDIC 本編として HDIC プロジェクトが公開する『篆隷万象名義』、『新撰字鏡』、観智院本『類聚名義抄』、宋本『玉篇』を収録、HDIC 拡張編として申請者が作成に参加したデータベースである『大字典』和訓データベース、夢梅本『和玉篇』データベース、『本草和名』データベース、『一切経音義』データベースを収録している。

第 7 章では、本論文をまとめ、成果として HDIC Viewer の利用の実績に関する統計を示した。