

博士学位請求論文審査要旨及び担当者

報告番号 甲 乙 第 号 氏名：木村 麻衣子

論文審査担当者

主査	慶應義塾大学文学部教授 文学研究科委員	谷口 祥一
副査	慶應義塾大学文学部教授 文学研究科委員	池谷 のぞみ
副査	慶應義塾大学名誉教授	田村 俊作
副査	国立情報学研究所名誉教授	宮澤 彰

論文題目：Name authority data and its model for non-Latin representations with special emphasis on Chinese characters

(漢字を中心とした非ローマ字表記による名称典拠データとそのモデル構築)

論文の要旨

木村麻衣子君による学位請求論文は、図書館目録の名称典拠コントロールにおいて、漢字文化圏に属する個人および団体に対する、漢字を中心とした非ローマ字表記による名称典拠データに焦点を当てた研究成果である。日本、中国大陸、香港、台湾、韓国の主要機関が作成する、自国または他国に属する個人・団体に対する名称典拠データの表記とデータ要素について現状を詳細に調査し、国際的な共有に当たっての問題点を綿密に整理している。併せて、国際的な共有の実現に向けて、非ローマ字言語圏の複雑な表記とその関連を表現可能にした典拠データモデルの構築、さらには構築したモデルに依拠した典拠データフォーマットの提案を行っている。

全体の章立ては、下記の通りである。

## 1. Introduction (はじめに)

1.1 Authority control in a global environment (国際環境における典拠コントロール)

1.2 Issues in non-Latin representations in global authority control (国際的な典拠コントロールにおける非ローマ字表記にかかわる問題点)

1.3 Purpose of the study (本研究の目的)

## 2. Literature review (先行研究レビュー)

2.1 History of rules and standards of authority control (典拠コントロールの規則と標準の歴史)

2.2 Authority control of names in non-Latin languages (非ローマ字言語の名称の典拠コントロール)

2.3 Writing systems in the Chinese character cultural sphere and its handlings in Western authority data (漢字文化圏の表記システムと欧米の典拠データにおける扱い)

2.4 Current conceptual models of authority data in non-Latin script languages and their

- inadequacies (非ローマ字言語の典拠データに対する既存概念モデルとその不足点)
3. Framework of name authority data (名称典拠データの枠組み)
    - 3.1 Characteristics of personal names in the Chinese character cultural sphere (漢字文化圏の個人名の特徴)
    - 3.2 Representations, data elements, and data structures (表記、データ要素、データ構造)
  4. Method and research objects (調査方法と調査対象)
    - 4.1 Method (調査方法)
    - 4.2 Current practices and policies of authority control in each area (調査対象地域の典拠コントロールの現況)
  5. Representations of Chinese name authority data in Chinese character cultures (漢字文化圏における中国人名・団体名典拠データの表記)
    - 5.1 Checkpoints and search terms (調査用チェックポイントと検索語)
    - 5.2 Adoption and character forms of Chinese characters (漢字の採用と文字種)
    - 5.3 Treatments and types of Romanization (ローマ字化の扱いと種類)
    - 5.4 Separation of surname and given name with a comma (コンマによる姓名の分かち)
    - 5.5 Representations in local languages outside China, Hong Kong, and Taiwan (中国大陸、香港、台湾以外での各国・各地域言語による表記)
    - 5.6 Discussion: differences in representations and problems to be solved (考察：表記の相違点のまとめと解決すべき問題点)
  6. Representations of Japanese name authority data in Chinese character cultures (漢字文化圏における日本人名・団体名典拠データの表記)
    - 6.1 Checkpoints and search terms (調査用チェックポイントと検索語)
    - 6.2 Adoption and character forms of Chinese characters (漢字の採用と文字種)
    - 6.3 Relating *yomi* to their corresponding *Kanji* (漢字とその読みとの対応づけ)
    - 6.4 Treatments and types of Romanization (ローマ字化の扱いと種類)
    - 6.5 Separation of surname and given name with a comma (コンマによる姓名の分かち)
    - 6.6 Representations in local languages outside Japan (日本以外での各国・各地域言語による表記)
    - 6.7 Names in *hiragana* (平がなによる名称)
    - 6.8 Discussion: differences in representations and problems to be solved (考察：表記の相違点のまとめと解決すべき問題点)
  7. Representations of Korean name authority data in Chinese character cultures (漢字文化圏における韓国人名・団体名典拠データの表記)
    - 7.1 Checkpoints and search terms (調査用チェックポイントと検索語)
    - 7.2 Adoptions of *Hangul* forms (ハングル形の採用)
    - 7.3 Adoptions of *Hanja* forms (漢字形の採用)
    - 7.4 Adoptions and types of Romanization (ローマ字化の採用と種類)
    - 7.5 Separation of surname and given name with a comma (コンマによる姓名の分かち)
    - 7.6 Representations in local languages outside Korea (韓国以外での各国・各地域言語による表記)

- 7.7 Discussion: differences in representations and problems to be solved (考察：表記の相違点のまとめと解決すべき問題点)
- 8. Representations of Vietnamese name authority data in Chinese character cultures (漢字文化圏におけるベトナム人名典拠データの表記)
  - 8.1 Research method (調査方法)
  - 8.2 Search results (検索調査結果)
  - 8.3 Discussion (考察)
- 9. Comparison of data elements of authority data in the Chinese character cultural sphere and the RDA framework (漢字文化圏の典拠データ要素の RDA との対照)
  - 9.1 Authority data elements in Japan (日本で採用されている典拠データ要素)
  - 9.2 Authority data elements in South Korea (韓国で採用されている典拠データ要素)
  - 9.3 Authority data elements in Mainland China, Taiwan, and Hong Kong (中国大陸、台湾、香港で採用されている典拠データ要素)
  - 9.4 A comparison to authority data elements in RDA (RDA の典拠データ要素との対照)
  - 9.5 Findings and recommendations (まとめと提言)
- 10. A proposal of a modification of FRAD model (FRAD 拡張モデルの提案)
  - 10.1 Differences among transliteration, transcription, and Romanization (翻字、翻音、ローマ字化の区分)
  - 10.2 The parent-child relationship between an original form and its Romanization and/or transcription (原語形とそのローマ字形および/または翻音形との親子関係)
  - 10.3 A modification of the FRAD model for recording names in non-Latin languages (非ローマ字言語による名称を記録するための FRAD モデルの拡張)
  - 10.4 Discussion (考察)
  - 10.5 Chapter conclusion (まとめ)
- 11. Trial on creation of new authority data formats (新たな典拠データフォーマット開発の試み)
  - 11.1 Modified MARC21/A (修正版 MARC21 典拠データフォーマット)
  - 11.2 The Resource Description Framework (RDF) (RDF 形式データフォーマット)
- 12. Conclusion (結論)

## 論文の概要

図書館目録における名称典拠コントロールとは、同一の個人や団体が多様な名称や表記の下に著者として出現する事象、あるいは異なる個人・団体が同一名称で出現する事象を適切に把握し、個人・団体の正確な同定と識別を図るという、データ品質管理の仕組みである。こうした典拠コントロールは近年、従来以上に重視される傾向にあり、欧米を中心に典拠データの国際的な共有が始まっている。

第 1 章では、国際目録原則覚書や「典拠データの機能要件」(Functional Requirements for Authority Data; FRAD)、Resource Description and Access (RDA) 等に見られる、近年の典拠コントロールを重視する国際的な動向を概観し、併せて VIAF (バーチャル国際典拠ファイル) など、さまざまな典拠データ共有プロジェクト

が実施されてきたことを確認している。しかしながら、国際的な典拠データ共有は主に欧米でなされているにとどまり、非ローマ字言語圏の名称の特徴が正しく理解され扱われていないこと、そのため非ローマ字言語圏に属する個人・団体に対する典拠データの共有は円滑に実現できないことを事例を挙げて指摘している。その上で、本論文の研究目的を、1)これまでその詳細が明らかにされてこなかった、漢字文化圏の主要機関と米国議会図書館が作成する名称典拠データにおける表記とデータ要素を明らかにすること、2)非ローマ字言語圏の複雑な表記とその関連を表現可能な典拠データモデルを構築すること、および3)構築したモデルを適用した典拠データフォーマットを提案することの3点に設定している。なお、本論文で扱う「名称典拠データ」は、個人と団体に対するものに限定し、地名や著作のタイトル（書名）等は含めないとしている。

第2章は、関連する先行研究のレビューを行い、まず典拠コントロールの全般的な歴史、漢字文化圏の表記システムと欧米の典拠データにおけるこれら表記の扱いについて跡づけている。次に、既存の典拠データモデルとしてのFRADや、MARC21典拠データフォーマット、RDA、DCMI (Dublin Core Metadata Initiative) Abstract Model等を取り上げ、いずれも複雑な非ローマ字言語の名称と表記を扱うには不十分であることを指摘している。

第3章は、漢字文化圏に属する中国人・日本人・韓国人・ベトナム人およびそれぞれの言語の団体名の特徴を概観した後、典拠データを「表記 (representations)」、「データ要素 (data elements)」、「構造 (data structures)」の3つの構成要素に分けるという枠組みを提示している。これまでの暗黙的な枠組みがデータ要素と構造に分けるものであったのに対して、先の課題解決に向けて新たに「表記」を設けている。

第4章では、研究方法と研究対象を述べている。対象機関への訪問によるインタビュー調査、各機関で採用されている目録規則、データフォーマット、各種マニュアル等の収集と参照、各機関で作成された典拠データの実例の収集と分析を組み合わせた詳細な調査を実施している。そして、それぞれの言語圏の名称典拠データにおいて特徴的と判断される事項を調査分析用の「チェックポイント」として設定し、それらに基づき実データや規則・マニュアル類を比較参照し、典拠データ共有に際して問題となりうる点を同定するという手順が説明されている。

第5章から第8章では、中国人名・団体名、日本人名・団体名、韓国人名・団体名、ベトナム人名のそれぞれについて、第4章で述べた研究方法に基づく調査と分析の結果が報告されている。1)中国人名・団体名を取り上げた第5章では、4つのチェックポイントが設けられ、調査の結果、漢字形の文字種が機関によって異なること、殆どの機関がローマ字形としてピンイン形を採用しているが、ウムラウトの扱いには違いが見られることなどが判明している。同様に、2)日本人名・団体名を対象とした第6章では、6つのチェックポイントを設けて調査し、漢字形に日本漢字を採用しているのは日本国内の機関に限定され、他の機関ではそれぞれの地域で通用している漢字が用いられていること、漢字形に対する読みは日本以外では採用されていないこと、ローマ字形はヘボン式ローマ字であっても規則の細部において異なったものであることなどを指摘している。3)韓国人名・団体名については（第7章）、5つのチェックポイ

ントを設定し、ハンダ形を必須扱いとするのは韓国国内の機関に限定されること、ローマ字化方式は韓国国内とその他の地域で方式が異なること、また韓国国内でも多様なローマ字化がなされていること、漢字形が不明な人名・団体名も多いことなどを明らかにしている。4)ベトナム人名については(第8章)、ベトナム国内で典拠データが作成されておらず、他国の機関が作成するベトナム人名典拠データを対象にした調査となり、漢字形とベトナム語形の両方を記録し、かつその関連を示しているのは1機関にとどまること、多くの機関はそのいずれかのみ記録している場合が大半であることなどを明らかにしている。

いずれの言語の人名・団体名の表記形も、単独では対象となる個人・団体を一意に識別し同定する要素としては不十分であり、いくつかの表記形を組み合わせることが有効かつ必要であることが第5章から第8章で確認されたことを受けて、第9章は、各機関が記録する名称表記以外の典拠データ要素を明らかにすること、そして最も包括的と考えられるRDAが規定するデータ要素と照合することを行っている。その結果、RDAのコア・エレメントについては殆どの機関で記録されており、それら以外に専攻分野、世系(特に日本)、性別、貫籍(特に中国)、団体の特徴、沿革等が多くの機関で記録されている実態を明らかにしている。併せて、いくつかの典拠データ要素を、典拠形アクセスポイント内の付記事項としてのみ記録し、独立した記録要素とはしていない国や地域(特に日本や中国)があることを指摘している。

第9章までの調査結果に基づき第10章において、典拠データ概念モデルとして国際的に認められているFRADモデルを、主に国際的な典拠データ共有の観点から名称表記の問題に対処できるよう拡張を試みている。具体的には、下記の3点についてモデルが反映できるよう拡張を行っている。

- 1)名称表記間の関連を表す「翻字(transliteration)」と「ローマ字化(Romanization)」がこれまで曖昧かつ混乱して用いられているため、ISO 5127:2001の定義に基づき、「翻字」、「翻音(transcription)」、「ローマ字化」の3つを区分する。
- 2)ローマ字化には、翻字によるもの、翻音によるもの、それら以外(英語圏で通用している「英語名」など)が混在しているため、図書館が典拠コントロールの目的でローマ字化した場合などには、適用したローマ字化の方式を明示する。
- 3)ある名称表記とそこから派生した表記との「親子関係」を認識し、それをペアとして明示する。たとえば、漢字等の非ローマ字言語による名称表記と、「非ローマ字による翻音」された表記や図書館が「ローマ字化」した表記の間には、こうした関係がある。

これら3点を反映させるようFRADモデルの拡張を試み、モデル内の実体「名称(Name)」と「統制形アクセスポイント(Controlled Access Point)」の属性群およびこれら実体間の関連群に必要な修正を行っている。

第11章では、第10章で提案した拡張FRADモデルに基づき、現在使用されている2つの典拠データフォーマットを拡張し、その妥当性を検証している。1)MARC21典拠データフォーマットについては、親子関係、非ローマ字による翻字、非ローマ字による翻音、ローマ字化という表記間の関連を示す関連識別コードを定義し、併せて関連識別コードならびにローマ字化方式の指示コードを入力するサブフィールドを新た

に定義した。2)RDF/XML についても、非ローマ字による翻字、非ローマ字による翻音、ローマ字化等の関連を示すことのできる新たなプロパティを定義し、これら表記間の関連を示せるようにした。いずれのフォーマットでも、拡張 FRAD モデルが反映できていることを確認し、併せてその課題を指摘している。

第 12 章では、研究成果を総括し、併せて提案を行っている。名称表記間の関連の記録や一定水準以上のデータ要素を含めた典拠データの作成がこれからも必要とされること、既存の典拠データの国際的共有には、VIAF 等のデータに対する人手による検証と修正が必要であることなどを指摘し、全体を締めくくっている。

### 審査の要旨

多種多様なデータがインターネット上に公開されるという状況が実現されつつある中で、品質の高いデータやメタデータの価値がますます認識されつつある。図書館が作成し管理する目録のメタデータもその一例であり、その品質を管理する主な仕組みの一つが名称典拠コントロールである。同一の個人や団体が多様な名称や表記の下に著者等として出現する事象、あるいは異なる個人・団体が同一名称で出現する事象を適切に把握し、個人・団体の正確な同定と識別を図り、それらが関与した情報資源を漏れなく集中化させるという仕組みである。この仕組みは、1) 個々の個人・団体に対して統一した名称すなわち典拠形を定め、それを一貫して適用する、2) 典拠形以外の名称は参照形として網羅的に典拠データにおいて、典拠形に加えて記録する、3) 個人の生没年や職業等、団体の創設年や場所など、同定と識別に有用な要素を併せて記録することから構成されている。こうした典拠データを如何にして一定水準以上に作成するか、それをどの範囲の情報資源に対する書誌データに適用するのかによって、目録全体の品質が左右される。

こうした典拠コントロールが近年、従来以上に重視されることに伴って、典拠データの国や地域を越えた国際的な共有が始まっている。個々の個人や団体に対して最も正確かつ詳細な典拠データは、その属する国・地域によって作成されており、それらの典拠データを共有または流用することによって全体として労力や経費の節減にも通じることがその背景にある。

しかしながら、これまでのところ国際的な典拠データ共有は主に欧米でなされているにとどまる。欧米で作成される典拠データは、名称の典拠形がローマ字表記であり、漢字等の非ローマ字表記は任意の参照形という扱いである。ところが、同一個人・団体が漢字等の原語形とローマ字表記の両者を持ち、かつ同姓同名という同一名称の多い漢字文化圏の名称にとって、原語形が十分に考慮されないことは大きな問題であり、實際上、共有に支障を来している。

一方、漢字文化圏に属する日本、中国大陸、香港、台湾、韓国では、同一個人・団体に対して相互に異なる多様な表記を含んだ典拠データをそれぞれ独自に作成しており、典拠データの共有も進められていない。漢字についても繁体字、簡体字、日本漢字があり、さらにカナ、ハングルなどがこれに加わる。また、原語形に対して、複数の翻字 (transliteration)、翻音 (transcription)、ローマ字化 (Romanization) の方式が併存しており、全体として複雑な状況を呈している。このため、国や地域を超え

た国際的な共有には、解決すべき課題が多数存在する。

本論文は、こうした漢字文化圏における名称典拠データ、すなわち漢字を中心とした非ローマ字表記による名称典拠データの現状を詳細に調査し、国際的共有に向けた問題点の整理と抽出、そして問題を解決する方策の提案を試みた成果である。これまで研究の必要性や重要性は認識されていたが、複数言語に精通する必要があるという言語の壁から本格的な取り組みはなされてこなかった研究課題に対して、本論文は真正面から取り組み、十分な成果を得ている。

本論文の第1の意義は、これまでその詳細が明らかにされてこなかった、漢字文化圏すなわち日本、中国大陸、香港、台湾、韓国の主要機関と米国議会図書館とにより作成されている、中国人名・団体名、日本人名・団体名、韓国人名・団体名、ベトナム人名のそれぞれの名称典拠データにおける表記の実態とデータ要素を詳細に明らかにしていること、その上で典拠データ共有に際しての問題点を整理し抽出したことである。

国内でも数通りの名称表記、国や地域をまたがったときにはさらに多様な表記となりうる漢字文化圏特有の状況下において、これまで各国内での表記法など、限定された範囲の調査と検討はなされてきたものの、漢字文化圏全体を対象にした研究は皆無であった。たとえば、日本人・団体に対する典拠データも作成国・地域によって異なる表記形式や文字種が用いられる。同様に、中国人・団体に対する典拠データも作成国・地域によって異なる表記形式や文字種が用いられている。こうした複雑な実態を、訪問によるインタビュー調査、各種マニュアルと実データとの照合など、多様な情報源を参照して広範に調査している。これだけの範囲を網羅しつつ詳細に調査した例は皆無である。多言語に精通してはじめてなした成果である。

最終的な成果として、漢字文化圏に属する個人・団体に対する典拠データの国際的共有を図る上での検討事項や問題点が網羅的に整理されており、今後の典拠データ共有に向けた基盤的な知見またはデータとなることは明らかであり、この点は高く評価される。

第2の意義は、前述の成果を踏まえ、国際的な共有の観点から特に重視すべき3点を取り上げ、典拠データ概念モデルとそれに依拠した典拠データフォーマットの拡張を試みていることである。その3点とは、1)これまで区分に曖昧さがあった名称表記間の関連を、「翻字」、「翻音」、「ローマ字化」に区分すべきこと、2)適用したローマ字化の方式(スキーム)が判明するものはそれを明示すべきこと、3)ある名称表記とそこから派生した表記との間には「親子関係」(依存・派生関係)を明示すべきことである。これら3点に絞って典拠データモデルおよび典拠データフォーマットの拡張を試みたこと自体、優れた着想であるといえよう。

まず、典拠データ概念モデルとして国際的に認められているFRADモデルについて、上記の3点を反映させるよう拡張している。その結果、非ローマ字言語圏の複雑な表記とその関連を表現可能な典拠データモデルの提示に成功している。漢字文化圏における表記法を踏まえた今後の議論や検討の基盤となるモデルが示されたものと評価できる。FRADを対象とした研究はこれまでなされてきているが、こうした観点からの拡張が示されたのは初めてであり、その点でも新規性をもつ研究成果である。

他方、難をいえば、FRAD 自体に内在する制約があり、多言語表記、そして表記と読みが分離している漢字文化圏の個人名・団体名にまつわる事項を素直に表現したモデルとは言いにくい部分もある。しかし、それは FRAD に代わる新たなモデルの抜本的な再構築にも通じる事項であり、本論文が取り上げた研究課題の範囲を大きく超えるものといえよう。

次に、拡張した FRAD モデルに依拠しつつ、実際に典拠データを記録し作成するためのデータフォーマットを提案し、その妥当性を検証している。あくまでも実際に使用されている典拠データフォーマットの拡張を意図し、MARC21 典拠データフォーマットおよび RDF/XML を拡張対象に採用している。その結果、いずれのデータフォーマットでも、拡張 FRAD モデルを反映し、前述の 3 点について表現可能であることが確認できている。

なお、RDF/XML については、国立国会図書館での採用方式にならい、SKOS (Simple Knowledge Organization System) などが定めた既存語彙をそのまま RDF プロパティとして流用した上で必要な拡張を施しているため、幾分制約の強い、不自然さを伴うフォーマットとなった印象が否めない。むしろ、意図する拡張に直接沿った新規のプロパティを独自に定義し適用した方が素直な拡張になったのかもしれない。ただし、この点は本論文の重要性を損ねるものでは決してない。

第 3 の意義は、第 2 章において展開された、典拠コントロールとそれに関連する研究およびプロジェクト等に対する包括的なレビューの有用性である。典拠コントロールの全般的な歴史、漢字文化圏の表記システムと欧米の典拠データにおけるこれら表記の扱い、そして既存の典拠データモデルと典拠データフォーマットという、本研究に関連する広範な事項への目配りの広さとレビュー構成の巧みさはそれ自体として評価に値する。

本論文は、6 編の国内外の査読付き学術雑誌に掲載された論文を核にしてまとめられた集大成である。さらに、そのうちの 1 編は、国内査読誌において優秀論文賞を受賞している。これらにより、国内外において既に高い評価を得ているといえよう。今後の典拠データの国際的な共有や円滑な相互運用を見据えたときに、本論文は重要な知見とデータを示す、極めて意義ある、オリジナリティの高い研究成果である。

以上の理由から、審査員一同は、本論文を博士（図書館・情報学）の学位授与にふさわしいものと判断する。