

1. 研究の背景と目的

近年、欧米を中心に典拠データの国際的な共有が始まっており、バーチャル国際典拠ファイル(VIAF)¹がその好例である。漢字文化圏の典拠データも欧米と共有されれば、検索の利便性が向上することは言うまでもない。しかし、現在のところ、欧米を代表する典拠ファイルであるLC/NACO典拠ファイルでは、典拠形アクセスポイントはローマ字表記であり、漢字等の現地語表記は任意の参照形という扱いである²。漢字文化圏の名前にとって、現地語形を十分に考慮されないことは大きな問題で、共有に支障を来すと考えられる。MARC21 Format for Authority Data (以下 MARC21/A) の Appendix C³では、現地語形を扱うための2つのモデルが示されている。しかし、これらのモデルが、漢字文化圏の著者名を表現するために十分なモデルであるのかどうかは、検討されなければならない。そのためには、漢字文化圏の著者名について、地域によってどのような特徴や相違点があるのかを明らかにしなければならない。

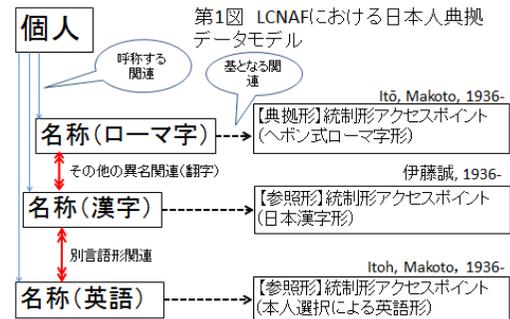
本研究の目的は、漢字文化圏の国々の著者名典拠データについて、典拠データの機能要件(以下 FRAD)⁴を応用した表記モデルを構築することで、それぞれの特徴と相違点を可視化し、分かりやすく提示することである。さらに、構築したモデルの活用例についても述べる。

2. 研究方法

漢字文化圏各地域の著者名典拠データの表記を、FRADの実体「名称」と「統制形アクセスポイント」およびそれらの間の関連を使用した実体関連モデルとして描いた。本研究では著者が現地語による姓名を1つだけ持ち、姓名に対する漢字やヨミはそれぞれ1つずつという想定でモデル化した。さらに、著者は英語で著作を為した場合に使用する英語名も持つものとする。各地域の人名使用の慣習は主に筆者のこれまでの研究成果に依ったが、中国の各地域で使用される英語名を確認するために、Web of

Science(以下 WoS)の検索結果を補助的に用いた⁵。

第1図は、LC/NACO Name Authority File (以下 LCNAF) での日本人、例として元東京大学教授の経済学者である伊藤誠の典拠データの表記を、FRADが定義している実体と関連のみを使用して表したものである。



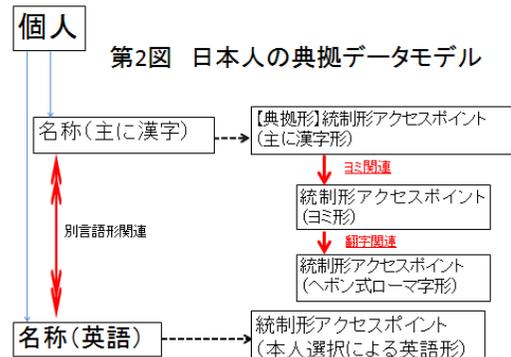
注:ヘボン式ローマ字形と、本人選択による英語形は同一の場合もある

個人が現地語と英語の名称のみを持つ想定であるにもかかわらず、ローマ字、漢字、英語の3つの名称を持つモデルとなったのは、FRADが統制形アクセスポイント間に翻字関連を定義しておらず、ヘボン式ローマ字形が日本漢字形から派生した統制形アクセスポイントであることを表すことができないためである。このように、FRADが現在定義している関連のみを使用してモデリングを行うと、実際には漢字形の翻字に過ぎないヘボン式ローマ字による名称を、あたかも個人が漢字の名称と同時に持っているかのようなモデルとなってしまう、実際の日本人の名称のあり方とのずれが生じる。そこで、以降のモデル図では、極力FRADに定義されている関連に従ったが、不足する場合は独自の関連を補うなどして、より各国・地域の著者名表記の実態に近いモデルを構築するよう努めた。なお、図が煩雑となるのを防ぐため、個人と名称の間の「呼称する関連」、名称と統制形アクセスポイントの間の「基となる関連」は、第1図以外では矢印のみを用いて表し関連名は明示していない。

3. CJK 各国・地域の著者名表記モデル

3.1 日本

日本人の名に使用できる漢字には制限があるが、その漢字をどのように読むかには制限がない。よって、同じ漢字を使用した姓名でも読み方は異なる可能性がある。作家が作品によって読み方を変えるケースはあるものの、ある人物の姓名の読み方は通常、1通りである。典拠データにおいて、読みはカタカナで記述される。読みはヘボン式ローマ字へ機械的に翻字させることができるが、例えば規則上“オウ”と“オオ”はともに“o”と翻字されるなどするため、ローマ字からヨミへの再変換は必ずしもできない。また、ヘボン式ローマ字形は、本人が選択した綴りによるローマ字形（英語形）とは異なる場合がある。以上のことから、日本人著者名の表記モデルを第2図のように描いた。

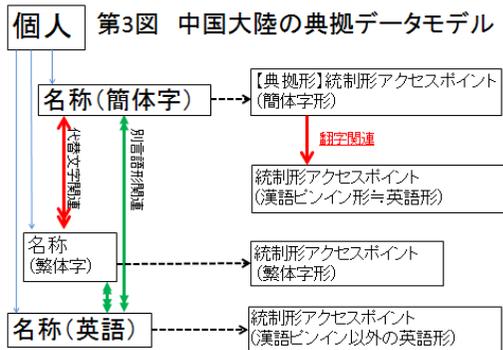


日本人名は、漢字だけでは成り立たず、ヨミだけでも成り立たないので、漢字形とヨミ形の間「ヨミ関連」を新たに定義した。ヨミ関連にあるもの同士は、著者の同定の目的のもとでは必ずセットで連鎖的に扱わなければならないものとする。ヘボン式ローマ字形はヨミから翻字したもので、両者の間を翻字関連（FRADとRDAでは名称間の関連とされる）で結んだ。ヨミからローマ字形への翻字は可能だが、ローマ字形からヨミへの正確な翻字はできない（不正確となる可能性がある）ため、この翻字関連は単方向のものである。

3.2 中国大陸

中国大陸では日常的に簡体字で姓名を表

記する。ピンインは漢字の普通話での発音を表したもので、1つの漢字に複数のピンインが存在する場合もあるが、多くの場合、漢字に対するピンインは1つに決まっている。中国大陸の著者は原則としてピンイン形を姓名の英語形として使用する⁶。以上のことから、中国大陸における著者名表記モデルを第3図のように描いた。

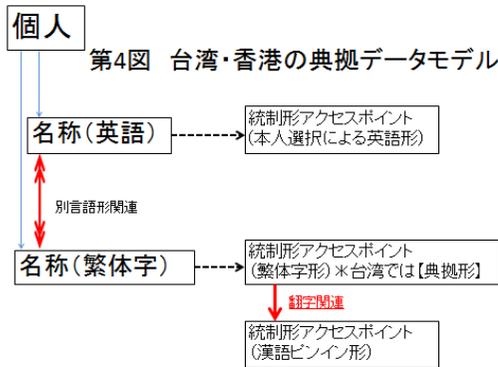


このモデルでは、個人が持つ簡体字の名称を典拠形とし、英語形でもある漢語ピンイン形と翻字関連で結ばれている。漢字からピンインへの翻字はできるが、逆はできないため、この翻字関連は単方向である。中国大陸の大学図書館コンソーシアムである中国高等教育文献保障系統(CALIS)の典拠データベースでは繁体字形の典拠形アクセスポイントも維持しているため、繁体字による名称もモデルに加えた。簡体字と繁体字の間には代替文字関連が存在するが、1:1では対応しない。なおFRADでは代替文字関連は統制形アクセスポイント間の関連とされているが、RDAでは名称間の関連として定義されている。

3.3 台湾

台湾では繁体字で姓名を表記するので、繁体字形が典拠形アクセスポイントとなる。台湾の人名のローマ字表記法は、大陸における漢語ピンインのように統一されていないことが、WoSで得た Taipeiの著者の英語名から確認された。従って、漢字形が分かればローマ字形が一意に定まるというものではない。中国大陸で使用される漢語ピンインは、台湾では日常的に使用されていないが、漢語ピンインが中国語をローマ字表記するための国際標準であるとの認識か

ら、典拠データでは漢語ピンインが必ず付与される。よって、台湾の著名典拠データモデルでは、漢語ピンイン形は繁体字形から派生したものととして翻字関連で結ばれる。漢語ピンイン以外のローマ字表記法には、以前標準的に使用されていたウェード式のほか、数種類が見られる。こうした表記を繁体字形から派生したものとも考えることもできるが、統一的な翻字ルールに則って表記されたものではないので、これらはむしろ、本人が複数のローマ字表記から自由に選択した英語名であると捉えたほうがよいだろう。英語名が、漢語ピンイン形と同一である場合もありうる。第4図に示す典拠データモデルは、台湾と香港に共通のものである。



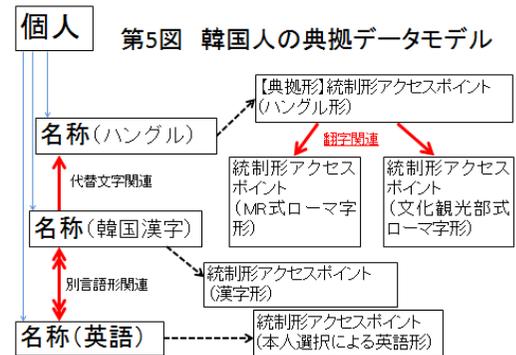
3.4 香港

香港の人々は「George」などの英語圏式のファーストネームを日常的に用いる⁷が、多くの場合本人の自称に過ぎず正式な名ではない⁸。また、それ以外の部分の姓名のローマ字表記法は統一されていない。香港の大学図書館コンソーシアム JULAC の人名典拠データベース HKCAN は、典拠形に LC 形を採用しており、多くの典拠形が漢語ピンイン形となっている。しかし、香港では普通話のローマ字表記である漢語ピンインは日常的に使用されていないため、香港の人々が名乗る英語名を重視したモデルを構築すべきである。第4図では、本人の自称する名や本人選択による広東語ローマ字表記は、合わせて本人の英語名とし、これと本人の繁体字形による名称の間を別言語形関連で結んでいる。漢語ピンイン形は、繁体字形から派生しているため、翻字関連

で結ばれている。なお香港はバイリンガル社会であるため、英語形と繁体字形のどちらを典拠形としてもよい。

3.5 韓国

韓国では日常的にハングルで姓名を記述する。多くの人は漢字による姓名があり、その漢字に当たるハングルの姓名を持つが、日本語のひらがなによる名前のように、当てはまる漢字のない名を持つ人もいる。複数の漢字が同じハングルに該当するため、ハングルによる姓名は同じでも、漢字形の姓名は異なる可能性がある。ハングル形から漢字形を派生させることはできないため、韓国人の典拠データモデル(第5図)では、代替文字関連が、単方向の関連として名称間にあるものとした。韓国人名のローマ字表記法は、統一されていないが、マッキューンライシャワー式(MR式)あるいは文化観光部式などの翻字法を用いて、ハングル形からローマ字形を機械的に作り出すことは可能である。ただし、翻字されたローマ字形から、もとのハングル形を完全に復元することはできない(不正確となる可能性がある)。よって、ハングル形とローマ字形の間には単方向の翻字関連があるものとした。実際の韓国人名のローマ字表記はこうした翻字法に必ずしも則っておらず、本人が複数のローマ字表記から自由に選択した英語名を持っていると考えられる。



4 モデルの活用例

統制形アクセスポイントの単方向の翻字関連として定義したローマ字形は、現地語を扱える利用者やシステムにとっては基本的に不要なものであると見ることができる。

例えば第5図のMR式ローマ字形は、

LCNAF で採用されている翻字法であるが、韓国国内では採用されていないし、文化観光部式ローマ字形も、人名についてはこれに従わなくてもよいことが明示されている。漢語ピンインは台湾、香港で用いられないばかりでなく、中国大陸でも、英語名として用いられる場合以外は、日常的に使用されているわけではない。日本人の場合、ヘボン式ローマ字形は NACSIS-CAT では採用されていない。このように、翻字関連にあるローマ字形は、典拠データの英語圏での流通のために便宜的に用意している統制形アクセスポイントであると言ってよい。このとき、単方向の矢印の先にある統制形アクセスポイントだけでは、現地語に正確に再変換することができず、同定識別も難しくなる。したがって矢印の反対側にある統制形アクセスポイントを必ず採用し、かつ翻字関連にある 2 つのアクセスポイントは、ヨミ関連と同じように、必ずセットで扱う必要がある。例えば、どちらかを付記事項として示すか、あるいは関連指示子で関連を示した上で、同定識別の際には 2 つのアクセスポイントの両方が一致するかどうかを見る、などである。このことは第 5 図の単方向の代替文字関連についても言える。例えば韓国の典拠データでは、漢字形はハングル形の付記事項としてアクセスポイントに記述されている。また、複数の国・地域で作成されたデータを同定識別する際に、単方向の矢印の先にあるアクセスポイントだけをマッチングさせるのではなく、矢印の根本にあるアクセスポイントの一致度を優先させることによって、精度の向上が期待できる。

漢字文化圏の著者名典拠データ共有のためには、本モデルで新たに定義したヨミ関連や、翻字関連の方向を考慮した典拠データフォーマットの開発が望ましい。今後はより複雑なケースでのモデル構築および、これらのモデルを包括的に扱うことのできるフォーマット作成を試みたい。

謝辞

本研究は、富士ゼロックス株式会社小林節太郎記念基金 2013 年度小林フェローシッ

プおよび平成 25 年度慶應義塾大学大学院博士課程学生研究支援プログラムより研究助成を受けて実施したものです。

注・引用文献

- 1) OCLC. "VIAF". <http://viaf.org/>, (accessed 2013-09-09).
- 2) Adam, Schiff. "Non-Latin script references in name authority records". Library of Congress. N ACO-Name Authority Cooperative Program of the PCC. 2009-06-01. <http://www.loc.gov/aba/pcc/naco/documents/Non-Latin%20Script%20References%20in%20Name%20Authority%20Records-RevJune2009.ppt>, (accessed 2013-09-09).
- 3) "Appendix C: Multiscript Records". MARC Standards. 2008-03-07. <http://www.loc.gov/marc/authority/ecadmulti.html>, (accessed 2013-09-09).
- 4) IFLA Working Group on Functional Requirements and Numbering of Authority Records. Functional requirements for authority data: a conceptual model. 2009., amended and collected in 2013, http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf, (accessed 2013-09-09).
- 5) 2013 年 7 月 24 日に、検索条件を CI(市)=Beijing かつ出版年=2013 として検索した、最新の上位 500 件の文献をダウンロードした。CI=Hong Kong, CI=Taipei でも同様にダウンロードした。そこから所属欄に各都市名を含む著者のみを抽出し、さらに重複した名を除いたところ、Beijing の著者 1384 名、Hong Kong の著者 1096 名、Taipei の著者 985 名を得た。
- 6) この事実を検証するために、WoS から得た Beijing の著者の姓名と漢語ピンインの音節が一致するかを確認した。姓名の一部がイニシャルによってのみ表記されている者 92 名、姓名ともに非中国系と見られる 2 名を除くと、漢語ピンインに沿わない英語名を持つ著者は 1290 名中 8 名(0.62%)であった。
- 7) WoS から得た Hong Kong の著者で、明らかに非中国系と思われる 31 名および名がイニシャルのみで構成される 186 名を除く 910 名のうち、323 名(35.5%)に英語圏式のファーストネームまたはミドルネームが見られた。
- 8) Li, David C.S. Borrowed identity: signaling involvement with a western name. *Journal of Pragmatics*. 1997, 28(4), p. 487-513.