

OPAC の FRBR 化を目指した人手による著作同定作業 FRBR 研究会の取り組み

谷口祥一（筑波大学）[†]、上田修一（慶應義塾大学） 横山幸雄（国立国会図書館） 鴨田拓哉（筑波大学大学院） 向當麻衣子（慶應義塾大学大学院） 宮田洋輔（慶應義塾大学大学院）
† taniguchi@slis.tsukuba.ac.jp

[抄録] OPAC の FRBR 化を図るには、個々の書誌レコードに対して著作同定処理が不可欠である。わが国の書誌レコードについて、これまで機械的な著作同定の試みはあるが、その性能には大きく限界がある。そこで、既存書誌レコードに対する主に人手による著作同定作業を、古典著作などに関して分担して進めている。本発表では、著作同定作業の概要とその結果の集計、採用した著作同定基準、作業結果の公開準備など、現時点の計画と作業状況等について報告する。

1.はじめに

OPAC の機能向上の方策の 1 つとして、FRBR（「書誌レコードの機能要件」）に依拠した著作に基づく集中化とナビゲート機能の実装が試みられている。こうした OPAC の FRBR 化を図るには、個々の書誌レコードに対して著作同定、すなわちいかなる著作に対応する体現形か判定することが不可欠である。著作の単位による検索または表示の機能を実現するには、著作の同定が不可欠となるからである。

これまでの試みは、書誌レコードを対象とした、著作の機械的同定（同一著作に属する候補の抽出とそれらに対する同一性の認定）によるものであった。OCLC の FictionFinder や WorldCat.org、オーストラリア国立図書館によるプロトタイプシステムなどがその代表例である。一方、著作に関する情報の記録が限定的なわが国のレコード群を用いた試行例には、JAPAN/MARC 書誌レコードを対象とした宮田や谷口による機械的同定の実験がある^{1),2)}。これらの試行結果によれば、機械的同定は有効ではあるが、大きく限界があるといわざるをえない。特に古典著作などについては機械的な著作同定は困難を極める。

そこで、発表者らは FRBR 研究会(仮称)を組織し、既存書誌レコードに対する主に人手による著作同定作業とその作業結果の公開準備を進めている。これは、著作同定は一度行われれば繰り返し行う必要はなく、また同定結果は複数機関で共有できることに依拠している。本発表では、現時点の計画と作業状況等について報告する。

なお、本研究と同様に、人手による著作同定を採用した事例に LibraryThing をあげることができる。ただし、明確な基準等がないまま同定作業を進めた結果、その質については疑問視されている。

2.著作同定作業手順

2.1 作業対象レコードの準備、判定結果の記録

現段階では、機械的同定が最も困難と想定されるわが国の古典著作（特に近世以前）を対象に同定作業を進めている。

JAPAN/MARC 書誌レコード（J-BISC レコード：明治期～2009.03 収録分）に対して、主要な古典著作ごとに J-BISC の検索機能を用いて漏れがない包括的な検索を行い、ヒットした書誌レコード群を JAPAN/MARC フォーマットでダウンロードし同定作業用ファイル（候補レコード群）としている。J-BISC における検索は、例えば著作「紫式部・源氏物語」の場合、「(タイトル : ゲンジモノガタリ or 源氏物語 or genji) or (件名 : 源氏物語) or (著者名[著者標目+責任表示] : 紫式部」という、ノイズを多く許容したものとしている。

次に、上記の候補レコード群に含まれる個々のレコードに対して、人手により当該著作に属するか否かを判定している。判定作業は分担して進めている。

個々の書誌レコードに対する判定結果は、該当する著作のタイトルが出現する箇所（以下、著作該当箇所）を記録する形式で蓄積している。具体的には、レコード識別番号（フィールド 001）の末尾に、著作の出現するフィールドの番号とサブフィールドコードを、00175008721-291A、00175001209-377A という形式で記録している。加えて、必要な場合には、該当するタイトルを、タグ<w></w>を用いてマークアップしている（図 1）。

なお、翻訳書の場合には、著作のタイトルは翻訳タイトル（25X、29X に記載されたもの）を採用している。

これらゆえ、単一の書誌レコードに対して、場合によっては複数の異なる著作を同定する

結果となる。この点は欧米における先行例とは異なり、網羅的な同定処理となるが、その分処理結果の集計や活用において複雑さを招くことになる。

また、当該著作に属すると認定しなかった書誌レコード、つまり異なる著作と判定した書誌レコードについても、その判定結果を記録している（たとえば、00175015629-D）。類似するが異なる著作と認定されたレコードを併せて蓄積することによって、著作の範囲の明確化を図り、判定結果の再考などに活用することを意図している。

00155002075-291A
020\$AJPSB55002075
100\$A19911219 1953 0JPN 1312
101\$AJPN
102\$AJP
251\$A 日本古典文学全集\$B 現代語訳\$D[第 11 卷]
270\$A 東京\$B 河出書房\$D1953
275\$A251p 図版\$B19cm
291\$A 枕草子\$F 池田龜鑑 訳
551\$A ニホン コソノ ブンガク センシュウ\$XNihon koten
bungaku zensyuu\$B251A1\$D11\$A ゲンダ 仁 や
SXGendaigoyaku\$B251B1
591\$A マクラノウシ\$XMakuranosousi\$B291A1
677\$A918\$V6
801\$AJPSBNational Diet Library,JAPAN
SC20031004\$GNCRTS2jpnmarc
905\$A918-G292

00155006562-377A
020\$AJPSB55006562
100\$A19920124 1965 0JPN 1312
101\$AJPN
102\$AJP
251\$A 日本文庫全集\$D 第 3
270\$A 東京\$B 河出書房新社\$D1965
275\$A475p 図版\$B20cm
291\$A 王朝日記隨筆集
350\$A 監修者：谷崎潤一郎等
377\$A 内容：土佐日記(池田弥三郎訳) 蜻蛉日記(室生犀星訳) 更級日記(井上靖訳) <w>枕草子</w>(田中澄江訳) 方丈記(佐藤春夫訳) 徒然草(佐藤春夫訳) 注釈(池田弥三郎) 年譜(阿部秋生) 解説(中村真一郎)
(以下略)

図1 著作「清少納言. 枕草子」の判定例

2.2 著作同定基準

著作同定作業を進めるに当たって、FRBRにおける著作の定義に依拠しつつ、必要な基準を順次設け、マニュアルにまとめている。現時点では、特にわが国の古典著作に合致した基準

整備を図っている。

まず、下記の基本の方針を採用している。

FRBR の示す基準に整合させる。「著作と著作の関連」に列記されている後継、補遺、追補、要約、改作、変形、模造の場合には、相互に異なる著作とする。他方、「表現形と表現形の関連」に列記されている縮約、改訂、翻訳、編曲の場合には、同一著作とする。

できるだけ既存の目録規則、その他基準類に整合させる。たとえば、古典著作の場合には、『国書総目録』『古典籍総合目録』を尊重し、できるだけ著作の単位を一致させる。

原則として書誌レコードのみを見て判定する。つまり、資料現物は基本的に参照しない。ただし、Web 上で参照可能なページ群 (Yahoo オークションなど) の参照は可とする。

国立国会図書館による作業方針とその作業結果をできるだけ活用して判定する。請求記号（分類記号）件名標目、著者標目の付与方針などを踏まえ、書誌レコードに示されている付与結果を尊重しつつ判定する。なお、同館における作業方針自体の変遷などにも配慮する。

基本の方針の上に、著作のタイプやカテゴリに合わせた基準を設けている。古典著作に関する基準として、下記のものを採用している。

同一著作とするもの：

- ・校注書
- ・現代語訳。部分訳を含む
- ・影印本
- ・縮約(abridgement) 要約(condensation)

異なる著作とするもの：

- ・評証書。原テキストが記載されている場合も含む
- ・学習参考書。原テキストが記載されている場合も含む
- ・児童書
- ・漫画
- ・索引
- ・書(原著作の本文を、ある書法をもって書き表したもの)
- ・ダイジェスト(digest) 抄録(abstract)
梗概・あらすじ

抜粋は同一著作とする。これは元のものと全体部分関連にあるが、便宜的に同一著作とする。たとえば、「源氏物語桐壺」は「源氏物語」と同一著作とする。「桐壺」、「帚木」、「空蝉」などをそれぞれ個別著作として（「源氏物語」とは異なる著作として）同定し、「源氏物語」との間に全体部分関連のリンクを形成するという方式はとらない。

3. 著作同定作業結果の集計

個別著作ごとの同定作業結果の一部を表1に示す。集計項目とその方法は次の通りである。

当該著作に属すると判定されたレコード数
同定作業対象レコード数

同定フィールドとサブフィールドの内訳

表1に示した以外にも、下記の項目について集計を行っている。

同定されたタイトル: フィールド 25X \$A(本タイトル) \$B(タイトル関連情報) 29X(多巻ものの各巻タイトル) \$A、\$B

同定された内容注記のタイトル: 377 \$A

責任表示: 25X \$F、29X \$F

著者標目: 751、79X

NDC 分類標目: 677 \$A。NDC の版次(\$V)ごとに集計する。

NDLC 分類標目: 685 \$A

NDLSH 件名標目: 650 \$B(個人名件名標目) 658 \$B(一般件名標目: 漢字形)。ただし、著作名による件名は、レコードが表す資料の主題として付与されたものである。よって、当該著作そのものの場合には、こうした件名は付与されていない。

出版者: 270 \$B

出版年: 100

なお、からは、25X、29Xにおいて著作が同定されたものを対象にした集計としている。内容注記(377 \$A)において同定された場合には、資料全体では当該著作とは相當に異なるものとなる場合が多いため集計の対象外とした。また、単一書誌レコードに複数の著作が記載されている場合には、25X、29Xにおける同定であっても、当該著作とは無関係の値が以降の集計結果に混じることもある。

加えて、J-BISCの検索機能を用いた検索結果集合と人手による同定結果との一致レコード数を測った。(a)著作の「著者名+タイトル」による検索、(b)タイトルによる検索、(c)著者名による検索、(d)件名による検索それぞれの検索結果に対して再現率と精度を測ったところ、著作「清少納言・枕草子」については、(a)再現率 0.496、精度 0.617、(b)0.707、0.172、(c)0.744、0.639、(d)0.030、0.023 となった。同様に、著作「清少納言・源氏物語」の場合、(a)再現率 0.565、精度 0.842、(b)0.988、0.268、(c)0.576、0.772、(d)0.020、0.013 となった。

これら集計結果および J-BISC 検索結果との照合から、単純な検索等によってこうした著作同定に代えられないこと、またプログラムに

よる機械的な同定処理は極めて困難であること、それゆえ人手による同定作業が不可欠であることが再確認された。

引き続き著作同定作業を実施する計画であるが、下記の課題について検討が求められている。

・判定結果の妥当性の検証。複数人によるダブルチェックまたは第3者による確認などの方策の必要性とその実施方法の検討。

・同定作業体制の拡充。作業担当者の増強など、作業体制の整備・拡充と必要なツールの開発。

・JAPAN/MARC 以外の書誌レコードに対する同定作業。JAPAN/MARC 書誌レコードに対する作業結果をできるだけ活用し、追加の手間を最小限に抑えた同定作業の方策の検討。

・未判定のレコード群に対する、比較的性能のよい機械的な同定法の開発。蓄積された同定作業結果を正解集合とした活用法の検討。

4. 著作同定作業結果の公開準備

同定作業結果をデータベースに蓄積し、順次公開していく計画である。現時点では2つの方式による公開、(a)OCLC Work Page を模した方式での作業結果公開と、(b)API による作業結果公開を計画し、システム試作を進めている。

(a)については、想定する使用者とその使用目的の検討がさらに必要であり、それによってシステムの機能仕様を確定させる。

(b)については、蓄積された著作同定結果をプログラムによって一括して(またはリアルタイムに)取得し、個々の OPAC に適用する活用法が想定される。それぞれの OPAC では、機械的な著作同定を全レコードに対して実行した上で、一部の著作(古典著作など)に対して API で取得した人手による著作同定結果を上書きし、機械的同定結果の修正を図るなどして活用することになる。

引用文献

- 1) 宮田洋輔「JAPAN/MARC レコードから自動構築可能な著作識別子の提案」『三田図書館・情報学会研究大会発表論文集 2008 年度』三田図書館・情報学会、2008, p. 69-72.
- 2) 谷口祥一、「FRBR OPAC 構築に向けた著作の機械的同定法の検証: JAPAN/MARC 書誌レコードによる実験」*Library and Information Science*. No.61, 2009, p.119-151.

表1 著作同定結果の集計

著作	古事記	竹取物語	伊勢物語	紀貫之 土佐日記	藤原道綱 母 蜻蛉日記	清少納言 枕草子	紫式部 源氏物語
著作に属するレコード数	149	126	134	114	60	133	1044
作業対象レコード数	2094	398	605	372	266	674	4112
著作同定箇所							
251A	106	77	74	74	29	94	948
251B	1		1				1
252A			6	1	5	3	
253A				1		1	
254A						1	
(多巻ものの各巻)							
291A	18	20	12	9	2	11	59
292A			5		5	1	
[小計(25X, 29X)]	[125]	[97]	[98]	[85]	[41]	[111]	[1008]
377A(内容注記)	24	29	36	29	19	22	36

著作	堤中納言 物語	菅原孝標 女 更級日記	今鏡	水鏡	西行 山家集	新古今和 歌集	源実朝 金槐和歌 集
著作に属するレコード数	80	77	32	30	42	111	34
作業対象レコード数	158	297	67	128	185	373	75
著作同定箇所							
251A	39	33	20	13	24	65	21
252A	3	3			1	3	
253A	2	3					1
254A		2					
(多巻ものの各巻)							
291A	8	4	2	2	7	22	3
292A	4	1		1			1
293A		2					
294A	1	2				1	
[小計(25X, 29X)]	[57]	[50]	[22]	[16]	[32]	[91]	[26]
377A(内容注記)	23	27	10	14	10	20	8

著作	宇治拾遺 物語	保元物語	平治物語	吉田兼好 徒然草	増鏡	松尾芭蕉 奥の細道	上田秋成 雨月物語
著作に属するレコード数	62	48	49	172	66	89	75
作業対象レコード数	170	86	111	977	157	997	349
著作同定箇所							
251A	30	27	26	115	35	63	43
252A	1		2	6	4	1	
253A				5	1		
254A				1			
(多巻ものの各巻)							
291A	18	5	4	18	6	1	8
292A	1		2	3	2		
293A		1		1			3
294A			1				
[小計(25X, 29X)]	[50]	[33]	[35]	[149]	[48]	[65]	[55]
377A(内容注記)	12	15	14	23	18	24	20