

# 著作同定における目録作成者間の一貫性

宮田洋輔 (慶應義塾大学) miyayo@slis.keio.ac.jp

## I. はじめに

1997年に国際図書館連盟(IFLA)が発表した『書誌レコードの機能要件』(FRBR)は図書館目録の概念モデルを提供した。FRBRモデルでは書誌レコードの対象となる実体を、「著作」、「表現形」、「体現形」、「個別資料」の4つの実体からなる階層的な実体として捉えている。

個別資料は手元の1つ1つの資料、体現形は具体的に資料として表れているものであるが、「個別の知的・芸術的創造」<sup>1)</sup>である著作と「著作の知的・芸術的表現」<sup>2)</sup>である表現形は物理的な形態をもたない抽象的な実体として定義されている。

図書館目録をFRBRに基づいて再構築することをFRBR化と呼ぶ。これまで作られてきた目録全体のFRBR化をおこなう際に、機械的なアルゴリズムに基づく方法と、人手による方法が考えられる。

機械的なアルゴリズムによる方法では既存の書誌レコードから自動的に著作に含まれる実体を特定するものである。OCLC<sup>3)</sup>や、日本では谷口<sup>3)</sup>などによって研究がおこなわれている。機械的なアルゴリズムによる場合、作業負荷は軽いものの、これまでに作成されてきた書誌レコード中に著作を識別するための要素が十分に表現されていなければ、実体の同定漏れが生じることがある。

一方、Batesは、これまで図書館が共同分担作業によって、総合目録を構築してきたように、図書館が協力し合っ<sup>4)</sup>てFRBR化をおこなっていくシナリオを提示した。手作業によるFRBR化は、これまでに作成されてきた書誌レコードの量を考えるとコストは嵩むものの、書誌レコード上には表現されておらず、機械的には判断不可能な事例にも対応できる可能性がある。

しかし、多くの目録作成者がかかわる状況で、著作と表現形という抽象的な実

体が扱われる際には、それぞれの作成者の間あるいは、目録の作成者と利用者との間で、その捉え方が異なる可能性が考えられる、

そこで、本研究では、目録作成者間での著作の捉え方の同一性、すなわち著作同定における一貫性に関する実験をおこなった。

## II. 研究の枠組み

FRBRの著作同定における一貫性を計測するにあたって、本研究では、索引作成者での索引語の一致の研究で用いられてきた枠組みを援用した。すなわち、索引作成者間の一貫性では、索引作成者が付与した索引語の集合の類似によって一貫性を算出していたところを、本研究では、著作をなす実体、つまり表現形と体現形とその間の関連の集合と対応付けた。

本研究では、著作( $W$ )を、表現形( $E_m$ )と体現形( $M_n$ )および表現形と体現形との間の関連( $R(E_m, M_n)$ )の集合としてとらえた。なお1つ1つの具体的な資料である個別資料は、本研究の枠組みからは除外した。本研究における著作は、以下のようにあらわすことができる。

$$W = \{E_1, E_2, \dots, E_m, M_1, M_2, \dots, M_n, R(E_n, M_n) \dots\}$$

たとえば、図1のように、目録作成者Aともう一人の目録作成者Bによる著作の同定結果があったとする。そのとき、2人の作成者による著作の同定結果は以下のように表現できる(ここでは関連をスペースの都合上に $R_1$ - $R_3$ と表記している)。

$$W = \begin{bmatrix} E_1 E_2 M_1 M_2 M_3 R_1 R_2 R_3 \\ 1 1 1 0 0 1 1 0 \\ 1 1 1 1 1 1 1 1 \end{bmatrix}$$

2名の作成者による著作の同定結果の一貫性を算出するために、索引者間一貫性の研究において用いられてきたHooperによる式を用いて算出した。

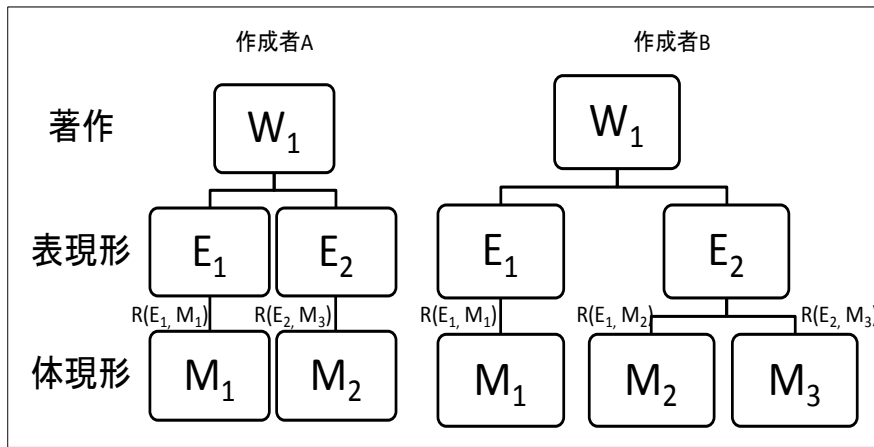


図 1 著作の同定結果の例

$$CP = \frac{A}{(A+M+N)}$$

ここで、CP は同定結果の一貫性を示す尺度である。A は 2 人の同定結果で共通して含まれていた実体の数、M と N は 2 人の作成者のどちらか一方にのみ含まれていた実体の数である。

図の例に基づいて、2 人の作成者間の一貫性を算出すると、A=5、M=0、N=3 で一貫性 CP は、以下のようになる。

$$\frac{5}{5+0+3} = 0.625$$

2 人の目録作成者によって同定された著作の集合がまったく同じ場合は 1 を、著作の集合に 1 つの実体の重なりもない場合は 0 を取る。一貫性が高ければ高いほど、2 人の目録作成者によって同定された著作の類似度は高い。

### III. 実験の概要

実験には、大学図書館などで実際に目録作成をおこなっている、あるいは最近までおこなっていた図書館員 7 名に参加を依頼した。参加者の平均目録作成年数

は 12 年である。また参加者はいずれも FRBR を知っており、関心を持っている。実験は 2009 年 7 月末から 9 月初め

にかけておこなった。

実験に用いた著作は、ある程度の大きさをもつ著作であることが望ましい。しかし、たとえば聖書や源氏物語のような古典の場合は、単純な目録に関する知識以上に、それらについての専門的な知識が求められるため、本実験では古典の大規模な著作は不相当であると考えた。

そこで、本研究ではそのような古典は対象にせず、ある程度の大きさが予測された、近現代以降のフィクション、ノンフィクションを対象著作として選んだ。対象著作は以下の 5 著作である。

1. 福澤諭吉『学問のすすめ』（ノンフィクション）
2. 芥川龍之介『羅生門』（フィクション）
3. George Orwell “1984”（フィクション）
4. Anthony Giddens “Sociology”（ノンフィクション）
5. Victor Fleming “Gone with the Wind”（フィクション、映画）

表 1 各著作の同定結果の概要

	学問のすすめ		羅生門		1984		Sociology		Gone with the wind	
	体現形	表現形	体現形	表現形	体現形	表現形	体現形	表現形	体現形	表現形
延べ同定数	326	105	689	257	290	88	107	74	81	28
異なり同定数	73	34	199	115	81	30	23	18	14	13
平均同定数	46.6	15.0	98.4	36.7	41.4	12.6	15.3	10.6	11.6	4.0
標準偏差	21.8	5.8	62.6	19.2	24.4	3.9	6.4	4.9	1.6	1.8

これらの著作に対して、NACSIS Webcat を用いて、それぞれの著作を構成する表現形を特定し、その表現形が属する表現形を特定することを依頼した。実体を特定した理由（書誌レコードのどの項目がカギになったか、参照した外部の情報源など）も記入を求めた。

表現形の識別子には、NACSIS Webcat での ID を用いた。表現形には統一的な識別子が存在しないため、参加者に FRBR の翻訳書を参考にそれぞれの言葉での記入を求めた。分析にあたって、不適当なデータの修正（余分な文字・スペースの削除、補助記号の除去）をしたうえで、表現形に付与されたそれぞれの名前を、著者が一意の識別子に対応付けた。

実験では、著作の同定には一定以上の時間を要すると考えたため、実験室におこなうような形にはせず、参加者に Excel スプレッドシートを配布し、記入を求める形式を取った。そのため、参加者の実験環境などは統制されていない。参加者は1週間から10日程度でシートへの記入を終え、ファイルを返送した。

#### IV. 実験結果

各著作に対して同定された実体数を表 1 に示した。“Gone with the wind”と“Sociology”は、表現形・表現形のいずれにおいても、比較的小さな標準偏差である。他方、羅生門では、表現形における同定数の標準偏差が62.6と参加者によってかなりの違いがあることがわかる。

上記の同定結果に対して、各著作に対して、「表現形」のみ、「表現形」のみ、表現形・表現形・その間の関連を含む「全体」の3つの水準で、各参加者対による一貫性を算出し、その平均を表 2 に示した。いずれの著作

でも全体での一貫性は50%以下で、低い一貫性を示した。

著作のタイプ（フィクション、ノンフィクション）によって、平均一貫性を表 3 に示した。2つの著作のグループで平均の差の検定をおこなったところ、表現形レベルでは99%水準で有意であった。

つぎに、それぞれの実体を特定した理由を集計した。理由も自由形式での記入であるため、著者がカテゴリ化し集計した。それぞれの著作と実体のレベルごとの合計での上位5位の特定理由と延べ同定数に占める割合を表 4 に示した。表現形レベルではタイトルが、表現形レベルでは責任表示がもっとも多かった。

表現形には統一的な識別子の命名規則

表 2 各著作に対する一貫性

		平均	標準偏差	最大値	最小値
学問のすすめ	表現形	50.4%	25.0%	96.9%	24.2%
	表現形	44.1%	14.6%	70.0%	25.0%
	全体	42.5%	20.0%	79.9%	17.3%
羅生門	表現形	35.8%	29.1%	88.8%	9.1%
	表現形	25.1%	17.3%	48.9%	1.6%
	全体	28.8%	22.5%	68.3%	4.3%
1984	表現形	41.7%	21.7%	81.4%	12.5%
	表現形	38.5%	6.9%	52.4%	25.0%
	全体	37.1%	17.3%	70.2%	13.2%
Sociology	表現形	58.0%	22.2%	100.0%	26.1%
	表現形	44.9%	22.4%	80.0%	11.1%
	全体	44.0%	18.4%	82.9%	15.8%
Gone with the wind	表現形	85.1%	15.3%	100.0%	61.5%
	表現形	32.9%	27.3%	80.0%	0.0%
	全体	48.4%	24.1%	90.0%	17.8%
5著作	表現形	54.1%	28.4%		
	表現形	37.1%	20.1%		
	全体	40.1%	21.3%		

表 3 著作のタイプ別による一貫性

		平均	標準偏差	最大値	最小値
フィクション	表現形	54.2%	31.5%	100.0%	9.1%
	表現形	32.2%	19.6%	80.0%	0.0%
	全体	38.1%	22.6%	90.0%	4.3%
ノンフィクション	表現形	54.2%	23.7%	100.0%	24.2%
	表現形	44.5%	18.7%	80.0%	11.1%
	全体	43.2%	19.0%	82.9%	15.8%

表 4 同定理由の集計

表現形	学問のすすめ		羅生門		1984		Sociology		Gone with the wind		合計	
	件数	同定数比	件数	同定数比	件数	同定数比	件数	同定数比	件数	同定数比	件数	同定数比
タイトル	243	74.5%	472	68.5%	224	77.2%	86	80.4%	68	84.0%	1,093	73.2%
責任表示	239	73.3%	471	68.4%	225	77.6%	86	80.4%	68	84.0%	1,089	72.9%
出版者	50	15.3%	159	23.1%	61	21.0%	13	12.1%	13	16.0%	296	19.8%
シリーズ名	41	12.6%	120	17.4%							161	10.8%
別タイトル	18	5.5%	91	13.2%	4	1.4%	7	6.5%	9	11.1%	129	8.6%
そのほか	95		246		101		14		32		488	
表現形												
責任表示	272	83.4%	643	93.3%	260	89.7%	86	80.4%	46	56.8%	1,307	87.5%
タイトル	284	87.1%	607	88.1%	229	79.0%	63	58.9%	56	69.1%	1,239	83.0%
内容注記	41	12.6%	91	13.2%	1	0.3%					133	8.9%
注記	21	6.4%	38	5.5%	6	2.1%	8	7.5%	56	69.1%	129	8.6%
内容細目	12	3.7%	81	11.8%					1	1.2%	94	6.3%
そのほか	65		165		101		99		26		456	

表 5 表現形実体数と記述との関係

	異なり実体数	記述数	記述/実体数比
学問のすすめ	34	98	2.9
羅生門	115	246	2.1
1984	30	83	2.8
Sociology	18	72	4.0
Gone with...	13	27	2.1

が存在せず、本実験では参加者に、報告書での記述を参考に自由記述での識別を依頼した。そこで、表現形の記述と識別される表現形の実体数の関係を表 5 に示した。表から、実際の実体数に対して、かなり多くの記述のバリエーションがあったことがわかる。

## V. 考察

全体の一貫性の 40.1%と、いずれの著作においても一貫性は高くはなかった。特に、表現形レベルでは 4 割を下回っており特に低かった。

FRBR を、これからの目録作成の基盤として考える場合には、一貫性のさらなる向上が必要である。そのためには、FRBR の概念の厳密化・明確化と、各実体を同定するための規則やガイドライン、また FRBR に基づいた書誌レコードの作成を支援するようなツールの開発が必要であろう。

## 謝辞

実験にご協力くださった参加者にここに心からの感謝を記します。

## 引用文献

- 1) 書誌レコードの機能要件：IFLA 書誌レコード機能要件研究グループ最終報告。和中幹雄，古川肇，永田治樹訳。本図書館協会，2004，121p.
- 2) Hickey, Thomas B.; Toves, Jenny. FRBR Work-Set Algorithm. [http://www.oclc.org/research/software/frbr/frbr\\_workset\\_algorithm.pdf](http://www.oclc.org/research/software/frbr/frbr_workset_algorithm.pdf), (2009-09-07).
- 3) 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. 2009, no. 61, p. 119-151.
- 4) Bates, Marcia J. Task force Recommendation 2.3 Research and Design Review: Improving User Access to Library Catalog and Portal Information: Final Report (Version 3). Metadata Enrichment Task Force, 2003, 58p. [http://www.loc.gov/catdir/bibcontrol/2.3\\_BatesReport6-03.doc.pdf](http://www.loc.gov/catdir/bibcontrol/2.3_BatesReport6-03.doc.pdf), (accessed 2009-09-07)