

<提言>

## 大学図書館におけるアジア資料の分類付与状況と、目録業務に関する提言

徳原 靖浩

### 1. はじめに

毎年、春になると、多くの大学図書館が新入生向けのガイダンスを開催する。図書館の基本的な利用法に加え、OPAC や CiNii での検索方法を教える機関も少なくないだろう。中には、検索結果から分類記号のリンクを辿って同じ主題の本を探したり、直接分類記号を入力して検索したりする方法を教えるところがあるかも知れない<sup>1)</sup>。

しかし、分類記号がすべての書誌データに付与されているわけでもなければ、付与されている記号も様々であり、要するに全ての図書が分類記号で検索できるわけではないことは、目録に携わる大学図書館職員ならば誰もが知っている事実である。では実際、分類検索ではどれほどの資料が検索結果からこぼれ落ちるのだろうか。日本語や英語の図書ならば、分類がなくても他の方法で見つけることができるかも知れないが、ではそれ以外の外国語の資料はどうだろうか。

特に、アジア研究で用いられる、アジアの諸言語で書かれた資料（以下、「アジア資料」とする）は、ラテン文字以外の文字で書かれているものも多く、日本語や英語よりも利用者にとって検索のハードルが高い。これらの図書を、和書と共通の分類記号によって検索することができれば、アジア研究を行う学生や研究者が資料を見つけやすくなる。ここでは、簡単な検索によってアジア資料の分類付与状況の概観をつかみ、どのような課題があるか考えてみたい。

### 2. CiNii Books の検索を使った分類付与率の検証

分類が付与されていないアジア資料の数を把握するため、CiNii Books の検索機能を使って、アジアの諸言語の書誌データの総数および、その内で日本十進分類法 (NDC)、デュイ十進分類法 (DDC)<sup>2)</sup>、米国議会図書館分類表 (LCC)、国際十進分類法 (UDC)、日本の国立国会図書館分類表 (NDLC) のいずれかの記号が付与されているもの、およびこれらのいずれも付与されていないものの数を調べた。CiNii Books 詳細検索画面による検索方法は以下の通りである。

1. 「検索対象」を「図書」に限定した上で、資料種別も「図書・雑誌」に限定（地図や楽譜が排除される）した。

---

<sup>1)</sup> 本稿では、主に書誌データに記入される分類記号を「分類」、この書誌分類による検索を「分類検索」とする。館毎に方式が異なる請求記号ないし書架分類から検索する方法は考察対象としていない。

<sup>2)</sup> デュイ十進分類法は一般的には DDC と略されるが、NACSIS-CAT における分類表コードは DC で表す。

2. 「分類」欄に、分類表のコード<sup>3)</sup>である「NDC\*」、「LCC\*」、「DC\*」、「UDC\*」、「NDLC\*」を一つずつ入力し、言語ごとに検索を行った<sup>4)</sup>。これらの記号のいずれも含まないもの（表中の「いずれもなし」）については、「NDC\* OR LCC\* OR DC\* OR UDC\* OR NDLC\*」でこれらの記号のいずれか1つでも含む書誌データの数を出し、この数と各言語の総数との差から算出した<sup>5)</sup>。

3. 「言語種別」欄には言語コードを一つずつ入力して検索した<sup>6)</sup>。こうすることで、書誌データ上でタイトルまたは本文に当該言語を含むものがヒットする<sup>7)</sup>。したがって、言語コードを複数持つ、二カ国語以上の言語で書かれた図書、例えばアラビア語日本語辞典はアラビア語と日本語の双方に計上されている。

4. 比較対象として、日本語、英語のほか、アジア研究で用いられることの多いロシア語、フランス語、ドイツ語、ハンガリー語などのヨーロッパ言語についても同様に算出した。

検索結果は表のとおりである。

表：アジア資料の分類付与状況

言語名 (言語コード)	地域	書誌データ 数	NDC あ り	LCC あり	DC あり	い ず れ も な し
日本語	東アジア	4,205,722	63%	0%	0%	36%
中国語	東アジア	656,371	14%	1%	1%	85%
朝鮮語	東アジア	119,196	9%	1%	12%	78%
モンゴル語	東アジア	8,996	13%	18%	2%	71%
チベット語	東アジア	7,348	9%	35%	8%	57%
ジャワ語	東南アジア	1,166	11%	38%	5%	57%
マライ語	東南アジア	7,419	31%	39%	11%	43%
タイ語	東南アジア	36,958	11%	25%	2%	68%

<sup>3)</sup> NACSIS-CAT で用いられる分類表のコードについては、「目録システムコーディングマニュアル (CAT2020 対応版) 付録 1.5 分類コード表」を参照されたい。

[http://catdoc.nii.ac.jp/MAN2/CM/furoku1\\_5.html](http://catdoc.nii.ac.jp/MAN2/CM/furoku1_5.html) (2021年4月13日最終閲覧)

<sup>4)</sup> コードにアスタリスクを付しているのは、前方一致検索のためである。

<sup>5)</sup> なお、上記の分類表の他に、NACSIS-CAT では中国科学院図書館図書分類法分類学 (CCAS)、中国図書館分類法分類学 (CLC)、韓国十進分類法 (KDC)、米国国立医学図書館分類表 (NLM)、ドイツ国立図書館分類表 (SG) の記号が付与された書誌データが少なからずあるが、CCAS (19,455 件) と CLC (42,762 件) はほぼ全て中国語、KDC (16,945 件) もほぼ全て朝鮮語、SG (99,640 件) は約9割がドイツ語図書である。また NLM (77,130 件) は大半が医学関係図書であると思われる。本稿の主旨に従いこれらは考察の対象から除外した。

<sup>6)</sup> 表中の言語名は「目録システムコーディングマニュアル (CAT2020 対応版) 付録 1.3 言語コード表」のものである。[https://catdoc.nii.ac.jp/MAN2/CM/furoku1\\_3.html](https://catdoc.nii.ac.jp/MAN2/CM/furoku1_3.html) (2021年4月13日最終閲覧)

<sup>7)</sup> 「CiNii Books・マニュアル・キーワードによる図書・雑誌検索方法」

[https://support.nii.ac.jp/ja/cib/manual\\_keyword](https://support.nii.ac.jp/ja/cib/manual_keyword) (2021年4月13日最終閲覧)

ビルマ語	東南アジア	6,680	55%	18%	1%	39%
ラオ語	東南アジア	2,447	59%	17%	2%	33%
インドネシア語	東南アジア	41,818	7%	33%	4%	61%
ベトナム語	東南アジア	14,781	9%	34%	3%	58%
タガログ語	東南アジア	2,096	16%	33%	10%	56%
ウルドゥー語	南アジア	22,949	12%	17%	2%	73%
梵語	南アジア	21,225	18%	39%	13%	49%
ベンガル語	南アジア	5,650	35%	36%	6%	45%
ヒンディー語	南アジア	33,319	38%	31%	5%	43%
ネパール語	南アジア	3,710	32%	48%	5%	40%
グジャラーティー語	南アジア	673	36%	39%	10%	37%
パンジャブ語	南アジア	1,157	71%	51%	14%	17%
ウイグル語	中央アジア	729	13%	26%	4%	63%
アルメニア語	中央アジア	697	5%	43%	18%	52%
タジク語	中央アジア	204	10%	36%	19%	48%
カザフ語	中央アジア	1,120	5%	27%	8%	64%
ウズベク語	中央アジア	1,150	6%	29%	10%	63%
オスマントルコ語	西アジア	4,458	6%	20%	2%	76%
アラビア語	西アジア	76,944	7%	24%	4%	70%
ペルシア語	西アジア	19,394	13%	24%	4%	65%
ヘブライ語	西アジア	5,979	3%	45%	31%	49%
トルコ語	西アジア	19,462	14%	35%	3%	54%
スワヒリ語	アフリカ	1,574	4%	51%	14%	43%
スウェーデン語	北ヨーロッパ	13,804	1%	29%	6%	68%
英語	西ヨーロッパ	3,809,126	6%	53%	50%	38%
ドイツ語	西ヨーロッパ	1,224,366	3%	11%	21%	74%
オランダ語	西ヨーロッパ	27,457	3%	6%	24%	72%
ラテン語	西ヨーロッパ	56,231	5%	15%	29%	65%
イタリア語	西ヨーロッパ	140,754	3%	17%	35%	61%
スペイン語	西ヨーロッパ	181,529	3%	25%	46%	49%
ポルトガル語	西ヨーロッパ	46,621	8%	19%	44%	49%
現代ギリシャ語	西ヨーロッパ	8,662	2%	11%	22%	74%
古代ギリシャ語	西ヨーロッパ	18,369	8%	26%	36%	54%
フランス語	西ヨーロッパ	707,289	4%	18%	27%	67%
ロシア語	東ヨーロッパ	273,596	2%	29%	9%	66%

ハンガリー語	東ヨーロッパ	8,929	4%	36%	7%	59%
合計						
1. 上記言語の合計		11,848,125	26%	22%	22%	48%
2. 中国語～トルコ語 の合計		1,124,096	14%	10%	3%	76%
3. NACSIS-CAT 書誌 データ総計		11,405,218	26%	23%	21%	48%

検索結果の最終確認日は2021年4月7日である。言語記号、分類記号ともに、一つの書誌データが複数の記号を付与されていることがあるため、検索対象となった書誌データの総和はNACSIS-CAT全体の書誌データ数を上回る点に留意されたい。

なお、UDC付与数は割合ではほぼゼロに近く、NDLCは日本語のみ26パーセントで他は0～2パーセントであるので表からは割愛した。

### 3. 検索結果に見られる分類付与状況とその傾向

言語によって、どの分類記号が最も多いか分かるよう、最も付与率の高い分類の数値を赤字にした。各言語を、それが主に用いられる地域別に並べてみると、一定の傾向があることがわかる。

東アジア言語では、日本語、中国語図書に付与されているのはNDCが主で、LCCの付与数はゼロに近い。意外なことに、日本語図書のNDC付与率は63パーセントしかない。日本語図書はNDLCの付与率が26パーセントあるが、「いずれもなし」が36パーセントであることから、NDLCの付与された書誌データのほぼ全てにNDCも付与されていると考えられる。「いずれもなし」の比率が最も高いのは、中国語の85パーセントで、こちらもほぼNDCに頼っていることが分かる<sup>8)</sup>。朝鮮語はNDCよりDCが多く、全体として付与率が低く見えるが、表にはないKDC(注5を参照)を付与されたものが16,563件(14パーセント)あり、最多である。KDCを考慮した場合は、朝鮮語の「いずれもなし」は75パーセントになる。

東アジアにも中央ユーラシアにも数え入れられることのあるモンゴル語、チベット語については、LCC付与数がNDC付与数を上回っている。また、東南アジアから西アジアにいたる地域のアジア諸言語も概ねLCCの付与率が高い。これは、USMARC<sup>9)</sup>から流用作成された書誌データが多く、流用元のデータのLCC記号を引き継いでいるためであろう。

<sup>8)</sup> ただし上の注5に示したように、CLCの付与されたものは中国語全体の6パーセント存在する。CLCも考慮した場合、「いずれもなし」は81パーセントになる。

<sup>9)</sup> 米国議会図書館(LC)のMARCレコード。NACSIS-CATでは、USMARCなどの各種MARC(機械可読目録)データを変換してデータベース内に取り込み、参照データセットとして格納している。これにより、目録作成者は自分でLCのデータベースを参照することなく、NACSIS-CATのシステム上でUSMARCを検索し、流用できるようになっている。

意外にも、一部の東南アジアと南アジアの言語、すなわちビルマ語、ラオ語、ベンガル語、ヒンディー語、パンジャブ語の NDC 付与率が他の言語に比して高いことが確認できる。母数が少ないということもあるが、これは言語の特徴に由来するというよりは、目録作成館の方針や作業者の資質によるところが大きいと推察する。

比較対象として統計をとって見たヨーロッパ諸言語との間には分かりやすい違いを見ることができる。英語、スウェーデン語、ロシア語、ハンガリー語は LCC の付与率が一番高いが、それ以外の西ヨーロッパ言語は DC 付与率が LCC を上回っている。現時点で筆者はこの理由を説明する材料をもたないが、日本の大学図書館において、洋書の請求記号に DDC を採用する機関が少なくないことも関係していると思われる。

表に示した言語全体で、上記の分類記号のいずれも付与されていないものは 48 パーセントに上ることがわかり、これは表にない言語を含めた NACSIS-CAT の図書データ全体を対象とした場合もほぼ同じ比率であった（表の合計 1 および 3）。本稿の主旨であるアジア資料に絞ってみると（表の合計 2）、いずれの分類も付与されていない率は 76 パーセントに跳ね上がる<sup>10</sup>。

#### 4. 分類付与状況から見えるもの

このデータからまず言えることは、CiNii Books に登録された書誌データのほぼ半分、そしてアジア資料に限れば 4 分の 3 は、分類記号からは検索できないということである。実際には、リンクされた親書誌データや子書誌データの分類から見つけられることもあると思われるが、概ねこのような状況にあると見て間違いではないだろう。

さらに言えば、比較的日本人に馴染みのある NDC の付与率だけで考えると、全体の 26 パーセント、アジア資料では 14 パーセントしか検索できないことになる。レファレンスカウンターで相談を受けた、NDC に精通した職員が検索したとしても、初動の検索で 7 割以上資料が見えていない状態にある。その意味でも、母数自体が多くないとはいえ、タイ語、ビルマ語、ヒンディー語、パンジャブ語の NDC 付与率が例外的に高いことは、地道な目録作成作業がもたらした意義深い成果と言える。

逆に、日本語図書の NDC 付与率がラオ語のそれと大差ないことや、欧米諸言語の NDC 付与率が総じて 10 パーセント未満であることは、目録作業が合理化・アウトソーシング化されていくなかで、分類作業が軽視されていることを物語っていると、筆者などは思うのだが邪推に過ぎるだろうか。

日本語図書については、株式会社図書館流通センターによる TRCMARC や国立国会図書館による JAPAN/MARC などの参照 MARC のデータが充実しているため、予め分類が付与されているケースが多いと考えられる。管見の限り、分類が付与されていない書誌には、科研費の報告書や展覧会カタログなどの灰色文献が多いようである。

---

<sup>10</sup> 日本語を除く。表中の中国語からトルコ語までが対象だが、スワヒリ語を加えても比率に変化はなかった。

試しに、CiNii Books でタイトルに「報告書」を含む、資料種別「図書・雑誌」、言語種別「日本語」の図書を検索すると、246,951 件がヒットし、その内で NDC の分類が付与されたものは 84,967 件 (34 パーセント) と、図書全体に比してかなり低い比率であることが分かる<sup>11)</sup>。

次に、同じ資料種別と言語種別で件名に「展覧会カタログ」を含むものを検索すると 60,496 件がヒットするが、その内 NDC が付与されたものは 22 件しかない。同様に、件名に「学位論文」を含むものは 164 件あり、NDC 付与率はゼロである。

この検索方法では、厳密に灰色文献を拾ったとは言えないが、商業出版物の分類は参照 MARC に頼り、灰色文献は分類しないという傾向が見て取れる。

従って、和書も含めて、分類付与率のバラつきは参照 MARC データに由来するものであり、図書館では請求記号付与のための書架分類は行うが、書誌に分類記号を追加することには消極的であると考えられる。

## 5. 目録・分類作業に関する提言

上の統計結果から明らかなように、OPAC や CiNii Books において、分類記号を直接入力したり、分類記号のリンクを辿って同じ主題の資料を探したりといった方法だけでは、主題検索の方法として不十分である。なお、主題検索には件名を使う方法もある。本稿では検証していないが、NACSIS-CAT では件名の記入も分類と同じく必須事項でなく<sup>12)</sup>、また複数の件名標目表が用いられるため、事情は分類と大きく違わないと見られる。

図書に分類を付与する作業は、基本的に情報源に書かれている通りに転記する項目 (タイトル、責任表示、出版事項など) や、ある程度機械的に抽出できる項目 (ページ数や大きさ、言語コードなど) と異なり<sup>13)</sup>、主題に関する知識や分類規則に関する知識、そしてそれを個別の資料に当てはめるための判断力、そして外国語資料であれば語学力も必要となる、専門的な知識と訓練を要する作業であって、その分、手間や時間や参考資料が必要となるが、だからといって省略してよいのだろうか。その場合、主題検索の方法はどのようなものが考えられるのだろうか。

分類記号・件名標目を直接検索する以外の主題検索については、様々な試みがなされている。図書館情報学の専門家でない筆者がここで辿々しい技術論を述べることは差し控えたいが、その多くは、分類や件名が付与されていることを前提にしていまいだろうか。

---

<sup>11)</sup> 2021 年 4 月 9 日検索。

<sup>12)</sup> NACSIS-CAT では、件名・分類記号の記入は必須ではなく選択事項である。ただし、「NACSIS-CAT/ILL 運用ガイドライン」(平成 23 年 1 月策定、平成 27 年 2 月改訂)には、件名・分類の付与について、「できる限り記入することが望まれます」とある。  
<https://www.nii.ac.jp/CAT-ILL/about/infocat/pdf/guideline.pdf> (2021 年 4 月 13 日最終閲覧)

<sup>13)</sup> 外国語資料、特にアジア資料の目録作成においては、これらの項目についても判断に迷うケースは多々ある。

例えば、そう遠くない未来に目録も分類も AI でできるようになる、とは図書館にいても時々耳にする言葉である。例えば『ニムロッド』のようなタイトルの本を読まずしてどうやって小説に分類できるのかは分からないが、AI がやるにしても、書誌情報だけでなく、ウェブ上の他の情報（芥川賞受賞作リストや、著者が小説家であることを示すプロフィールなど）を参照する必要があるのではないか。だとすれば、図書館員がやらなくても、結局は別のところで誰かが主題分析を行う必要があるのではないだろうか。

従来の分類法に拘泥せずとも、オントロジーのようなセマンティック・ウェブに向けた情報組織化に、分類研究の現代的な発展としての可能性を見いだす意見もある<sup>14)</sup>。しかし、今日明日にも分類法や件名法が取って代わられる、ということではなさそうであるし、むしろ分類が付与されているほうがセマンティック・ウェブにとっても好都合であると思われる。

より現実的な方法として、分類記号や異なる件名標目を同定したりマッピングしたりする様々な方法が提案、実証されている<sup>15)</sup>。実用化される可能性の高い研究だと思われるが、やはり分類か件名のいずれかが付与されている資料が対象となる。

筆者自身、これらの技術によって主題検索が発展することに期待している。しかし、図書整理の現場ではただその発展を待つて楽観視することはできない。これらの研究によって技術的可能性が示されても、図書館システムへの実装の段階まで面倒を見てくれる人間がいなければ、その恩恵を受ける順番は図書館員にまで回ってこないだろう。

そして何よりも、巨大な集積された情報を扱う情報学では、分類も件名もなく、電子化もされていない外国語資料の存在が見落とされ、技術革新から置いて行かれてしまう懸念がある。そうならないためにも、少なくとも分類か件名のどちらか一つでも付与することは必要ではないだろうか。

書誌データの遡及的な登録・修正作業については、量が多いがためにあたかも不可能であるように言われることが多いが、分類が付与されていない 500 万件の書誌データに分類を付与することは現実的でないように見えても、アジア資料に関しては、言語ごとのデータの母数はそれほど多くないので、各言語の専門家が一人いれば、書誌データの整備状況を把握し、適切な手当てをしていくことは決して難しいことではない。請求記号に主題分類を含んでいる館であれば、さらにハードルは下がるだろう。たとえ少しずつでも分類の付与率を上げることは、リンクト・データやセマンティック・ウェブ技術によって書誌データを統合的

---

<sup>14)</sup> Hjørland, Burger. 2002. "[Review] The Future of Classification. Rita Marcella and Arthur Maltby. Hampshire, England: Gower Publishing; 2000; 144 pp. Price \$99.95 (ISBN: 0-56607992-5)." *Journal of the American Society for Information Sciences and Technology* 53(1): 57.

<sup>15)</sup> 石田栄美 2005 「日本十進分類法と基本件名標目の相互マッピングの試み」『文化情報学』12(1): 1-11. ; 谷口祥一・木村麻衣子 2017 「機械学習によって NDLSH 細目付き件名標目に対する NDC 代表分類記号を同定する試み」『日本図書館情報学会誌』64(2): 60-76. ; 村上幸二 2016 「NDC と参照のリンクを用いた小学校件名標目表と基本件名標目表 (BSH) の統合的検索手法の検討」『日本図書館情報学会誌』62(3): 181-199.のほか、これらが紹介する先行研究を参照されたい。

に利用していく上でも大きなメリットとなるはずである。

さらに、外国語に長けたカタログラーを長期的に雇用し、目録作成や分類付与だけでなく、レファレンスやパスファインダー、文献リスト、解題書誌といった、従来からある情報サービスにも協力してもらうことで、分類や件名の不備を補う形で資料提供をすることができると。書誌データに不備があったとしても、選書や整理に関わった人間はどんな図書があるか把握しているものである。

こうした多言語資料の整理技術や資料組織化に長けた人材を、各大学に1名ずつでも、あるいは大学図書館界全体で10数名でも長期的に確保し、多言語資料の状況に関して様々な知見を共有していくことは、全文テキストがあればAIで主題分類できるというような話よりも現実的であり、利用者にとってのメリットも大きい。学術研究や教育の本来の目的とは関係ない「雇い止め」の論理によって、多言語資料の組織化に携わる人材が失われることは、資料の死蔵や、学生の図書環境の悪化などにつながり、特に外国語資料を抱える大学図書館にとって大きな機会損失となりうる。このことを大学図書館関係者に改めて認識していただくべく、敢えて時代に逆行する提言をさせていただいた次第である。

(とくはら やすひろ 東京大学附属図書館アジア研究図書館上廣倫理財団寄付研究部門  
(U-PARL))

2021年4月25日受理