

NDLSHにおけるNDC代表分類記号と書誌レコードの分類記号はどの程度一致するのか

谷口 祥一[†]

[†]慶應義塾大学文学部
taniguchi@z2.keio.jp

尾形 沙由美[‡]

[‡]聖マリアンナ医科大学医学情報センター
gt_sym@marianna-u.ac.jp

[抄録] NDLSHにおいて、件名標目に対して概念上で対応するNDC新訂9版の分類記号が「代表分類記号」として示されている。本研究は、国立国会図書館作成の書誌レコード(1997年から2014年3月分まで)に付与されたNDLSH件名標目とNDC分類記号の組み合わせの中で、代表分類記号が付与されている割合を調査し、その分析を試みた。同様の調査を行った先行研究を踏まえつつ、普通件名と地名件名に分けた集計、分類記号の前方一致を加えた集計、レコード内先頭出現の件名標目に限定した集計など、詳細な集計と分析を行った。

1. はじめに

国立国会図書館件名標目表(NDLSH)に登録された、ある範囲の件名標目には、概念上で対応する日本十進分類法(NDC)新訂9版の分類記号が示されている。これらの分類記号は「代表分類記号」と呼ばれている。たとえば、件名標目「メタデータ」は、「014」(資料の収集、整理、保管)と「014.3」(目録法;記述目録法)という代表分類記号をもつ。

このような個々の件名標目と分類記号の概念間の対応づけと、個別の書誌レコードに付与された件名標目と分類記号の組み合わせとが、実際にどの程度一致するのかは興味深い事柄である。集計結果に表された実態を踏まえた上で、これら2種類のデータをさらに有効活用する方策の検討が可能になると考えられる。

実際に書誌レコードに付与された件名標目と分類記号の組み合わせの中で、代表分類記号が付与されている割合を調べた先行研究に、和田らによる研究がある¹⁾。本研究は、彼らの研究を踏まえつつ、そこでは採用されていない、a)件名標目の種別を普通件名と地名件名に分けた集計、b)レコード内で先頭に出現する件名標目に限定した集計、c)細目を除去した主標目による照合と集計、d)分類記号の階層構造に沿った前方一致の追加、e)一致率の平均として、算術平均に加え、加重平均の適用などを試みる。

2. 普通件名の出現状況と代表分類記号の有無

国立国会図書館がNDC新訂9版を採用した1997年以降、2014年3月分までの書誌レコードをJ-BISC、JM-BISCから全件取得し、それらのうちNDC9版の分類記号とNDLSHの件名標目(MARCフィールド658:一般件名標目)の両方をもつレコードを抽出した。抽出した件名標目に対して、WebNDLAuthorities

のAPIを用いて、件名標目の種別(普通件名、地名件名などの種別)を判別した。

普通件名の出現状況を表1に示した。件名標目の異なり単位を指す「件名単位」、件名標目と分類記号の組み合わせの異なりを指す「件名・分類記号ペア単位」、ペア出現ごとの単位である「ペア出現単位」、件名標目が付与されたレコードの単位を指す「レコード単位」のそれぞれにおいて集計した。

普通件名の異なり数は78,346、分類記号とのペア異なり数は202,802となった。普通件名が出現したレコードにおいては、平均1.37個の普通件名によるペアが付与されている。また、件名標目当たり平均2.59(SD 5.89)の異なる分類記号と組み合わせられている。件名標目当たりの平均ペア出現回数は12.91(SD 69.47)、件名・分類記号のペア当たり平均出現回数は4.99(SD 50.38)であった。

表1に、件名標目における細目の有無、代表分類記号の有無とのクロス集計結果を併せて示した。件名単位で見た場合、18.4%は細目をもたない、主標目のみからなる件名標目であり、残り81.6%は細目付き件名(例:「資料組織化 || 論文集」)であった。それに対して、ペア出現回数で見たときには、66.7%が細目のない件名で、残り33.3%が細目付き件名となる。つまり、細目をもたない件名は、異なり数では少ないが、数多く付与され出現するという状況を表している。件名・分類記号のペア単位で見たときには、細目の有無は半々である。

代表分類記号をもつ件名標目は全体の18.3%に当たる14,345個あり、残り81.7%はそれをもたない。また、代表分類記号をもつ件名標目の殆どは細目のない件名であり、逆に代表分類記号をもたない件名標目の殆どは細目付きという実態であった。

3. 書誌レコードの分類記号と代表分類記号との一致の程度

個々の件名標目ごとに、書誌レコードに付与された分類記号と当該件名標目をもつ代表分類記号とを照合し、完全一致、前方一致、もしくは不一致と判定する。複数の代表分類記号をもつ件名標目の場合、いずれかの代表分類記号に合致すれば一致とみなす。また、前方一致は、書誌レコードの分類記号が代表分類記号の低位区分であったときを一致に含めている。

件名単位では、代表分類記号をもつ 14,345 件名のうち、完全一致は 12,098 件名 (84.3%)、前方一致では 12,455 件名 (86.8%) が一致となった (表 2)。件名単位の場合、8 割以上のものが代表分類記号と完全一致または前方一致となる分類記号を少なくとも 1 個以上もつことを示している。件名・分類記号ペア単位では、13.7%が完全一致であり、前方一致に拡張しても 20.7%の一致にとどまった。一致とならないペアが全体の 8 割近く存在するという、代表分類記号以外の多様な分類記号との組み合わせが多数発生している状況にある。ペア出現単位では、56.7%が完全一致、前方一致を含めると 66.7%が一致となり、出現した半数以上の場合において代表分類記号と一致していることが分かる。ペア出現単位における代表分類記号と一致する割合の高さは、NDLSH において個々の件名標目に代表分類記号を割り当てていることの有用性をも示すものと考えられる。たとえ、書誌レコードへの件名標目と分類記号の付与が代表分類記号とは無関係に行われていたとしても、事後的な一致割合の高さは意味あるものであろう。

次に、件名ごとの一致率の平均を求めると、まず算術平均では、完全一致で 54.94、前方一致を加えたとき 60.25 となった (表 3)。ここでは、代表分類記号と一致する分類記号が出現しないという不一致の件名すなわち一致率 0 の件名を加えて計算している。件名・分類記号ペア単位の場合、一致率の算術平均は、完全一致 7.68、前方一致を加えても 8.16 となった。前述したペア単位全体での一致数の低さを反映した値である。

一致率の加重平均を求めると、件名単位の場合、完全一致で 39.92、前方一致を含めて 51.32 となった。ペア単位の場合でも、加重平均はそれぞれ 37.08、47.74 であった。算術平均の場合に比べて大きく値は改善しており、完全一致や前方一致となるペアに出現回数が比較的集中している状況が読み取れる。

1) レコード内先頭出現ペアに限定した場合

複数の件名標目を単一資料に付与する場合、対象資料の主な主題を表す件名標目から順次付与される。そこで、主な主題に相当する件名標目に限定するため、複数の件名標目がフィールド 658 に記録されている場合、その先頭が普通件名であるときのみ当該件名標目を抽出し先の照合を行った (表 2, 3)。

件名単位では 9 割以上が完全一致そして前方一致となり、先頭の件名標目に限定しないときに比べて微増となった。それに対して、ペア単位では対象となるペア数が 32,514 へと大幅に減少し、それに反比例して完全一致が 41.2%、前方一致を含めて 55.5%と、大幅に一致の割合が上昇している。ペア出現単位においても、相当程度に一致の割合が増加している。また、一致数の増加を受けて平均一致率も大きく上昇し、算術平均の場合、件名単位では 76.46 (完全一致)、82.25 (前方一致)、ペア単位では 30.31 (完全一致)、31.79 (前方一致) となった。加重平均についても、その上昇幅は大きい。代表分類記号との一致の程度は総じて高いと結論づけられる。

2) 主標目レベルに統合して照合した場合

細目を除去し主標目のレベルで統合した集計を併せて行った (表 2, 3)。その結果、件名単位、ペア単位、ペア出現単位のいずれにおいても 99%以上が代表分類記号をもつことになり、細目付き件名のままでは除外されていた多数の事例を包含させた形で照合することができる。ただし、細目が付加されたときには主標目とは異なる分類記号が代表分類記号として割り当てられるべきである場合も相当数あることに留意しなければならない。

照合の結果、件名単位での一致数は微増にとどまった。ペア単位では完全一致数、前方一致を含めた一致数は、細目付き件名標目のまま扱ったときに比べ相当数の増加を見せたが、代表分類記号をもつ全体数も増加しており、割合で見るときには顕著な増加とはなっていない。ペア出現単位、レコード単位についても状況は同じである。

平均一致率についても同様な状況にあり、細目付き件名標目のまま扱ったときに比べて、むしろ全体的に低下している。細目付き件名標目は、主標目に割り当てられている代表分類記号とは異なる (かつ前方一致ともならない) 分類記号が付与されている事例も多いことを表している。これより、細目を単純に除去し主標目の代表分類記号をもって細目付き件名標目の

表 1. 普通件名の出現状況

	件名単位		件名・分類記号ペア単位		ペア出現単位		レコード単位	
総数	78,346		202,802		1,011,135		740,196	
細目なし	14,394	(18.4%)	102,683	(50.6%)	674,632	(66.7%)		
細目あり	63,952	(81.6%)	100,119	(49.4%)	336,503	(33.3%)		
代表分類記号あり	14,345	(18.3%)	102,647	(50.6%)	675,642	(66.8%)	516,828	(69.8%)
細目なし	14,323	(18.3%)	102,427	(50.5%)	674,114	(66.7%)		
細目あり	22	(0.0%)	220	(0.1%)	1,528	(0.2%)		
代表分類記号なし	64,001	(81.7%)	100,155	(49.4%)	335,493	(33.2%)	223,368	(30.2%)

表 2. 普通件名における一致数

	件名単位		件名・分類記号ペア単位		ペア出現単位		レコード単位	
代表分類記号あり	14,345	(18.3%)	102,647	(50.6%)	675,642	(66.8%)	516,828	(69.8%)
完全一致数	12,098	(84.3%)	14,088	(13.7%)	383,277	(56.7%)	370,228	(71.6%)
前方一致数	12,455	(86.8%)	21,263	(20.7%)	450,598	(66.7%)	428,447	(82.9%)
不一致数	1,890	(13.2%)	81,384	(79.3%)	225,044	(33.3%)	88,381	(17.1%)
レコード内先頭出現に限定照合	12,887		32,514		458,760		458,760	
完全一致数	11,668	(90.5%)	13,386	(41.2%)	352,664	(76.9%)	352,664	(76.9%)
前方一致数	11,947	(92.7%)	18,050	(55.5%)	404,204	(88.1%)	404,204	(88.1%)
不一致数	940	(7.3%)	14,464	(44.5%)	54,556	(11.9%)	54,556	(11.9%)
主標目レベル照合	14,866		142,526		1,009,048		739,245	
完全一致数	12,636	(85.0%)	14,782	(10.4%)	495,929	(49.1%)	475,005	(64.3%)
前方一致数	13,047	(87.8%)	31,219	(21.9%)	665,325	(65.9%)	622,609	(84.2%)
不一致数	1,819	(12.2%)	111,307	(78.1%)	343,723	(34.1%)	116,636	(15.8%)

表 3. 普通件名における平均一致率

	件名単位		レコード内先頭出現		主標目レベル	
	件名単位	ペア単位	件名単位	ペア単位	件名単位	ペア単位
算術平均						
完全一致	54.94	7.68	76.46	30.31	53.79	5.61
前方一致	60.25	8.16	82.25	31.79	60.35	6.08
加重平均						
完全一致	39.92	37.08	66.81	62.36	33.56	31.19
前方一致	51.32	47.74	82.03	76.82	50.62	47.21

代表分類記号とすることは適切ではないといえよう。

4. 地名件名の出現状況と代表分類記号との一致の程度

141,722 レコードに地名件名が出現し、168,985 回、分類記号とのペアとして出現している (表 4)。レコード平均では 1.19 ペアの出現となる。なお、件名出現単位では 168,982 回の出現であり、ペア単位の出現と変わらない。

件名単位では 18,860 個の異なる地名がこれらレコード群に出現した。そのうち、細目をもたない地名が 20.1%、細目をもつものが 79.9%

であった。また、代表分類記号をもつ地名は 6.0%にすぎず、大多数は代表分類記号をもたない。細目がない地名であっても、そのうち 2,859 個には代表分類記号が示されていない。普通件名以上に、代表分類記号が割り当てられていない状況を表している。

地名による件名標目について、その一致数を表 5 に、一致率の平均を表 6 に示した。普通件名の場合と比べてみると、地名件名については、全体的に一致する割合が低く、前方一致によっても一致数の大幅な増加は見られない。

レコード内で先頭出現した件名標目が地名であるものに限定したときには、代表分類記号

表 4. 地名件名の出現状況

	件名単位		件名・分類記号ペア単位		ペア出現単位		レコード単位	
	件数	割合	件数	割合	件数	割合	件数	割合
総数	18,860		38,019		168,985		141,722	
細目なし	3,782	(20.1%)	28,838	(75.9%)	32,891	(19.5%)		
細目あり	15,078	(79.9%)	9,181	(24.1%)	136,094	(80.5%)		
代表分類記号あり	1,124	(6.0%)	4,354	(11.5%)	17,663	(10.5%)	16,499	(11.6%)
細目なし	923	(4.9%)	3,295	(8.7%)	9,063	(5.4%)		
細目あり	201	(1.1%)	1,059	(2.8%)	8,600	(5.1%)		
代表分類記号なし	17,736	(94.0%)	33,665	(88.5%)	151,322	(89.5%)	125,223	(88.4%)

表 5. 地名件名における一致数

	件名単位		件名・分類記号ペア単位		ペア出現単位		レコード単位	
	件数	割合	件数	割合	件数	割合	件数	割合
代表分類記号あり	1,124		4,354		17,663		16,499	
完全一致数	613	(54.5%)	625	(14.4%)	7,417	(42.0%)	7,337	(44.5%)
前方一致数	624	(55.5%)	687	(15.8%)	7,665	(43.4%)	7,565	(45.9%)
不一致数	500	(44.5%)	3,667	(84.2%)	9,998	(56.6%)	8,934	(54.1%)
レコード内先頭出現に限定照合	1,002		3,029		13,073		13,073	
完全一致数	577	(57.6%)	587	(19.4%)	6,031	(46.1%)	6,031	(46.1%)
前方一致数	582	(58.1%)	629	(20.8%)	6,237	(47.7%)	6,237	(47.7%)
不一致数	420	(41.9%)	2,400	(79.2%)	6,836	(52.3%)	6,836	(52.3%)
主標目レベル照合	923		9,983		79,936		70,597	
完全一致数	520	(56.3%)	534	(5.3%)	3,881	(4.9%)	3,730	(5.3%)
前方一致数	547	(59.3%)	804	(8.1%)	9,424	(11.8%)	8,869	(12.6%)
不一致数	376	(40.7%)	9,179	(91.9%)	70,512	(88.2%)	61,728	(87.4%)

表 6. 地名件名における平均一致率

	件名単位		レコード内先頭出現		主標目レベル	
	件数	割合	件数	割合	件数	割合
算術平均						
完全一致	32.45	8.38	38.65	12.79	20.76	1.92
前方一致	33.61	8.52	39.63	12.98	23.96	2.11
加重平均						
完全一致	27.98	27.88	35.54	35.43	1.58	1.55
前方一致	29.56	29.37	37.59	37.38	3.09	2.89

をもつ件名標目数、ペア異なり数、ペア出現回数
のいずれも総数としては減少している。それ
に対して、完全一致、前方一致の占める割合は、
限定しない場合よりも、多少とも上昇を見せて
いる。ただし、普通件名の場合のような大幅な
上昇ではない。

また、主標目レベルにまとめた上での照合で
は、代表分類記号をもつ件名標目数は減少し、
ペア異なり数とペア出現回数は大きく上昇す
る。主標目に統合した分だけ、代表分類記号を
もつペア異なり数とペア出現数が増加する結
果となった。ただし、ペア単位とペア出現単位

における完全一致数、そして前方一致を含めた
一致数の割合はともに低い。主標目（地名）に
対応づけられた代表分類記号では、細目が付加
されたときの分類記号を表しえない、つまり細
目（「地名のものの細目」）が付いたときには細
目を反映した、異なる分類記号が割り当てられ
るべきであることを表しているといえよう。

注

1) 和田匡路, 川向直樹「件名標目表の代表分類
記号と書誌レコードの一致する割合」第 58 回日
本図書館情報学会研究大会発表要綱. 2010,
p.61-64.