

Title	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
Sub Title	To what extent are the representative NDC classification numbers for subject headings in NDLSH and NDC numbers assigned to bibliographic records matched?
Author	谷口, 祥一(Taniguchi, Shoichi) 尾形, 沙由美(Ogata, Sayumi)
Publisher	三田図書館・情報学会
Publication year	2016
Jtitle	Library and information science No.75 (2016.), p.37- 66
Abstract	<p>【目的】国立国会図書館件名標目表(NDLSH)において、細目をもたない普通件名など、ある範囲の件名標目に対して、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「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、NDLSHが示す代表分類記号が付与されている数および割合を調査する。</p> <p>【方法】1997年から2014年3月分までの書誌レコードから、NDC新訂9版分類記号とNDLSH件名標目の組み合わせをすべて抽出し、それらを普通件名と地名による件名に分けて集計した。また、各件名標目の代表分類記号は、Web NDL Authoritiesを用いて取得した。</p> <p>【結果】1) 普通件名の場合、代表分類記号をもつ件名標目は全体の20%以下にとどまり、その殆どは細目をもたない。書誌レコードに付与された分類記号との照合では、件名単位で80%以上、件名・分類記号ペア単位で10%強が代表分類記号と完全一致となった。分類記号の前方一致まで含めたときには、件名単位でもペア単位でも一致となるものが増加した。対象をレコード内先頭出現の件名標目に限定したとき、一致数および平均一致率の大幅な上昇が見られた。それに対して、細目を除去し主標目のレベルですべての件名標目を照合したところ、これらの値はむしろ低下した。2) 分類記号の側から集計したときには、代表件名標目をもつ分類記号は30%強であった。書誌レコードに付与された件名標目との照合では、分類記号単位で90%弱、分類記号・件名ペア単位では10%強が完全一致となった。継承一致という独自の照合を導入すると、一致となるものが相当数増加した。3) 代表分類記号をもつ地名件名は僅か6%にとどまり、件名・分類記号のペア単位でも12%にとどまる。これらのうち、完全一致となるのは、件名単位で50%強、ペア単位で15%程度であった。</p> <p>Purpose : NDLSH are lists of subject headings, some of which have their corresponding classification numbers in NDC 9th ed., and are called "representative classification numbers." This paper investigates the extent to which the representative numbers in NDLSH match the classification numbers assigned to bibliographic records. It reports the detailed results on their matching, while referring to a preceding research on the same topic.</p> <p>Methods : Pairs of NDC numbers and NDLSH subject headings were extracted from bibliographic records created between 1997 and 2014. General topical subject headings and geographic ones were examined separately. The representative classification numbers for individual subject headings were obtained from Web NDL Authorities.</p>

	Results : 1) For general topical headings, less than 20% of them have representative NDC numbers, most of which do not have sub-headings. More than 80% of subject headings with representative NDC numbers have records with NDC numbers matching the representative ones, while more than 10% of the NDLSH-NDC combinations that occurred in records were in agreement with those shown in NDLSH. Left-hand match on classification numbers increased the number of matching cases. Limiting the NDLSH-NDC pairs to those that occurred first in each record brought a substantial increase in the number of matching cases. 2) Only 6% of geographic subject headings have representative NDC numbers. More than 50% of such headings have records with NDC numbers that match the representative ones, while 15% of the NDLSH-NDC combinations occurred in records that agreed with those shown in NDLSH.
Notes	原著論文
Genre	Journal Article
URL	http://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00003152-00000075-0037

原著論文

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

To What Extent Are the Representative NDC Classification Numbers for Subject Headings in NDLSH and NDC Numbers Assigned to Bibliographic Records Matched?

谷口 祥一
Shoichi TANIGUCHI

尾形 沙由美
Sayumi OGATA

Résumé

Purpose: NDLSH are lists of subject headings, some of which have their corresponding classification numbers in NDC 9th ed., and are called “representative classification numbers.” This paper investigates the extent to which the representative numbers in NDLSH match the classification numbers assigned to bibliographic records. It reports the detailed results on their matching, while referring to a preceding research on the same topic.

Methods: Pairs of NDC numbers and NDLSH subject headings were extracted from bibliographic records created between 1997 and 2014. General topical subject headings and geographic ones were examined separately. The representative classification numbers for individual subject headings were obtained from Web NDL Authorities.

Results: 1) For general topical headings, less than 20% of them have representative NDC numbers, most of which do not have sub-headings. More than 80% of subject headings with representative NDC numbers have records with NDC numbers matching the representative ones, while more than 10% of the NDLSH-NDC combinations that occurred in records were in agreement with those shown in NDLSH. Left-hand match on classification numbers increased the number of matching cases. Limiting the NDLSH-NDC pairs to those that occurred first in each record brought a substantial increase in the number of matching cases. 2) Only 6% of geographic subject headings have representative NDC numbers. More than 50% of such headings have records with

谷口祥一：慶應義塾大学文学部，港区三田 2-15-45

Shoichi TANIGUCHI: School of Library and Information Science, Keio University, 2-15-45 Mita, Minato-ku, Tokyo

e-mail: 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，川崎市宮前区菅生 2-16-1

Sayumi OGATA: Medical Information Center, St. Marianna University School of Medicine, 2-16-1 Sugao, Miyamae-ku, Kawasaki-shi, Kanagawa

e-mail: gt_sym@marianna-u.ac.jp

受付日：2015年5月28日 受理日：2016年1月17日

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

NDC numbers that match the representative ones, while 15% of the NDLSH-NDC combinations occurred in records that agreed with those shown in NDLSH.

- I. はじめに
- II. 先行研究
- III. 件名標目の出現状況と代表分類記号の有無
 - A. 取得データ
 - B. 普通件名の出現状況と代表分類記号の有無
 - C. 分類記号ベースの普通件名の出現状況
 - D. 地名件名の出現状況と代表分類記号の有無
- IV. 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
 - A. 一致数と一致率の平均
 - B. 普通件名についての一致の程度
 - C. 分類記号ベースの普通件名についての一致の程度
 - D. 地名件名についての一致の程度
- V. 考察
 - A. 相関係数による出現状況の考察
 - B. 不一致ペアおよび代表分類記号をもたないペアからの有効なペアの抽出
- VI. おわりに

I. はじめに

国立国会図書館件名標目表（以下、NDLSH）に登録された件名標目のうち、細目をもたない普通件名や一部の地名件名などには、概念上で対応する日本十進分類法（NDC）新訂9版の分類記号および国立国会図書館分類表（NDLC）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」と呼ばれており、“その件名標目を使用される場合に一般的に書誌に付与される分類記号を示す。しかし、これは書誌に付与する分類を制限するものではない”¹⁾[p. 75]と説明されている。たとえば、件名標目「メタデータ」にはNDC新訂9版の分類記号「014」と「014.3」が代表分類記号として示されている。NDC分類記号「014」は「資料の収集、整理、保管」、同「014.3」は「目録法（記述目録法）」を指す項目である。同様に、件名標目「資料組織化」も代表分類記号「014」をもつ。この対応関係を逆転させ分類記号の側から見れば、「014」には件名標目「メタデータ」と「資料組織化」が対応し、

「014.3」には「メタデータ」が対応することになる。本稿では、分類記号に対応するこれら件名標目を「代表件名標目」と仮に呼ぶ。

a) このような件名標目表と分類法（分類表）の直接的な対応づけ、すなわち個々の件名標目と分類記号との概念間の対応づけと、b) 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、すなわち対象となる資料の主題表現におけるそれらの共起とが、実際にどの程度一致するのかは興味深い事柄である。ある件名標目が書誌レコードに出現したとき、分類記号にその代表分類記号が付与されることが殆どまたは大半であるのか、あるいはそれとは異なる分類記号が付与されることが大半であるのかは、以下の2つの点から興味深い。

①件名標目表と分類法の対応づけデータ、およびこれら両者を用いて個別書誌レコードへの件名標目と分類記号の付与がなされたデータという2つが揃っていること自体極めて限られており、一致の程度に関する集計結果は得がたい存在である。個別資料の主題表現結果としての分類記号と

件名標目の組み合わせと、代表分類記号・代表件名標目という概念間の対応づけとが、殆どもしくは大半において一致するのであれば、分類法と件名法という特性の相違がここでは表出しないことになる。他方、大半において一致しないというのであれば、それは対象資料のトピックを包括的に表現する分類項目を採用するというNDCおよびその適用法と、対象資料の主たるトピックに加えて副次的なトピックも表現するというNDLSHおよびその適用法との、両者の特性の相違がもたらす帰結であるのか、あるいはそれ以外の理由によるものかなど、さらなる究明を必要とする。最終的には個別の分類法・件名標目表およびそれらの適用方針に依存する部分も多く、一般化には慎重であるべきとはいえ、上記の集計結果は主題表現方式に関わる有用な知見と考えられる。

②先の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。仮に、書誌レコードの分類記号として代表分類記号が付与されることが殆どであるならば、分類記号の付与作業においてこの代表分類記号は大いに参考になるはずであり、それを活用した分類作業支援システムも有効に機能するであろう。分類記号から件名標目への対応づけを活用した件名作業支援においても同様である。あるいは、現行のNDLSHの拡充、たとえば代表分類記号とは不一致となる分類記号の中から有用な組み合わせを抽出しNDLSHに加えるとの課題に、実態から浮かび上がった規則性を適用し活用することなどが考えられる。それに対して、代表分類記号とは異なる分類記号が付与されることが大半であるならば、そうした状況を正しく理解し、利用者にも適切に知らせることが必要となろう。たとえば、OPAC等の検索システムに件名標目表と分類表の対応づけを組み入れて利用者に提供する場合に、a) これらの対応づけ結果と個別書誌レコードへの件名標目と分類記号の付与結果という組み合わせとの相違を、さらにはb) それら概念的に対応する件名標目と分類記号のそれぞれを用いた検索結果が相互に大きく異なる集合となることを、利用者にもどのように説明し

理解を得るのかという問題を解決しなければならない。逆にいえば、書誌レコードの分類記号として代表分類記号と一致するものが付与されることが殆どであるならば、この問題は重大視する必要がない。

上記のような問題意識を恐らくは共有して、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、その件名標目に対応する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²⁾。彼らは、a) 国立国会図書館作成の書誌レコード(1997年から2009年3月)に出現したNDLSH件名標目とNDC新訂9版の分類記号、NDLCの分類記号を抽出し、他方b) Web NDL Authorities(国立国会図書館典拠データ検索・提供サービス)から取得したNDLSH一括ダウンロード用ファイル(件名標目とその代表分類記号が含まれる)とを照合し、c) 書誌レコードに出現した分類記号と代表分類記号との一致率を件名標目ベースに集計し、併せてd) 出現した件名標目と代表件名標目との一致率を分類記号をベースに調査している。彼らの成果によって概況は判明したが、学会発表という制約もあり、細部に渡る報告とはなっていない。

この和田らの研究に触発された本研究は、彼らの成果を踏まえつつ、さらに展開した詳細な集計と分析を実行し、詳細に実態の解明を図ることを目的としている。本研究の目的は、以下の2つである。

①書誌レコードへのNDLSH件名標目とNDC分類記号の付与状況(出現状況)の詳細な集計: 付与の実態を多角的に表し把握するため、複数の集計の単位を設定し、それぞれの集計値を求める。ただし、集計する範囲は、代表分類記号と書誌レコードの分類記号との一致の程度を計る範囲に限定する。

②代表分類記号と書誌レコード付与の分類記号の一致数および一致率の算出、代表件名標目と書誌レコード付与の件名標目の一致数および一致率の算出: 和田らによる先行研究で一部の概数が示されているが、その値は筆者らの直感的な予想よ

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

りも低い。照合の範囲(条件)を変化させたときに、それらの値がどのように変化するのか、最大でどの程度の一致が得られるのか集計と分析を行う。併せて、一致の程度を測る複数の指標を導入し多角的な把握を試みる。

これらの研究目的に沿って、具体的には、下記のような、先行研究では採用されていない集計と分析を試みる。a) 件名標目の種別を普通件名と地名による件名に分けた集計と分析を行う、b) 書誌レコードには当該資料の主な主題を表す件名標目から順番に付与されるため、レコード内で先頭に出現する件名標目に限定した集計を行う、c) 数の上で大きな割合を占める細目付き件名標目には殆ど代表分類記号が付与されていないため、細目を除去した主標目による照合など、細目付き件名を考慮した集計と分析を行う、d) 照合においては、分類記号の階層構造に沿った一致(前方一致、継承による一致)を加える、e) 照合結果を示す一致率の平均は、算術平均に加え、加重平均を適用するなど新たに試みる。上記a)からd)は照合範囲を変化させることに該当し、e)は一致の程度を測る複数の指標の導入に該当する。

なお、本研究ではNDC分類記号のみを対象とし、NDLC分類記号は扱わない。これはわが国での分類法の採用状況に依拠したにすぎず、NDLC分類記号についても同様の集計と分析が可能である。

II. 先行研究

先に取り上げた和田らによる研究以外で、本研究に直接関わる研究例は見当たらない。彼らの研究および本研究では、a) 件名標目表と分類法との対応づけ、b) 個別書誌レコードに出現した件名標目と分類記号の組み合わせという2種類のデータを用いている。こうした研究が可能となるのは、特定の件名標目表と分類法の全体同士で対応づけが人手によって大規模になされていることが条件となる。しかしながら、このような大規模な対応づけは米国議会図書館件名標目表(LCSH)と米国議会図書館分類表(LCC)の対

応づけなど、実施例は極めて限られており、それゆえ上記のような研究例はその数が限定される。

2種類のデータのうちの1つ、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を用いた研究には、石田によるものがある³⁾。石田はNDCと基本件名標目表(BSH)の書誌レコードへの付与事例データを用い、分類記号と件名標目の自動付与について実験を行った。代表的な機械学習手法であるサポートベクタマシン(SVM)と独自の相対出現率を用いた自動付与実験を行っている。さらに先行する研究としてFrankらは、LCSHによる件名標目からLCCの分類記号を推定する実験を、SVMを用いて行っている⁴⁾。

もう一方のデータである件名標目表と分類表の全体同士での対応づけ、すなわち人手による対応づけ結果を活用した研究は複数ある。たとえば、NDC、BSH、NDLSHという複数の主題表現方式を組み合わせる検索システムの研究がある。村上らは、NDCをベースにBSHとNDLSHとをリンクさせ検索を支援するシステムを開発している⁵⁾。さらに、村上らはBSHとNDCの両者を組み入れた検索システムSubject Worldを開発している⁶⁾⁻⁸⁾。また、松本らは、NDCとNDLSHの両者を検索に用い、利用者を適切なNDC分類記号に導く検索システムNDC Finderを試作している⁹⁾。現在、WebNDL AuthoritiesはNDLSH件名標目自体の検索に加え、NDC、NDLCの分類記号から対応する件名標目を検索できる機能を実装している。

また、谷口はBSHとNDLSHの件名標目に割り当てられているNDC代表分類記号を手がかりに、BSHとNDCの組み合わせ、NDLSHとNDCの組み合わせにおいて、どの程度上位下位関係の階層構造が相互に一致するのか定量的に調査している¹⁰⁾。

他方、特定の分類法と件名標目表との間の事後的な対応づけを図る研究、特に機械的な照合により対応づけを図る研究に目を転ずれば、そこには多数の研究事例が存在する。Chanらは複数回にわたり、それら研究について包括的なレビューを行っている^{11),12)}。特に大規模な実験とその成果

の公開を進めている事例には、OCLCの研究チームによるものがある^{13),14)}。機械的な照合には多様な方法が試みられており、書誌レコードに共に付与された実績という共起の情報に基づくものなどがある。

分類法と件名標目表という範囲を超えてオントロジーと呼ばれる領域を見てみると、そこには複数の概念体系、語彙体系、分類体系などの間の写像（マッピング）や対応づけ（アラインメント、マッチング）に関する膨大な研究例が存在する^{15)~17)}。それらはいずれも人手による写像形成や対応づけをもたない体系間において、多様なアルゴリズムにより体系間の事後的な写像形成や対応づけを試みる研究である。したがって、書誌レコードに付与された件名標目と分類記号の共起データから件名標目表と分類法の体系間の対応関係を導き出すようなアプローチも、ここに含まれる。

III. 件名標目の出現状況と 代表分類記号の有無

A. 取得データ

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^{18)~21)}とJM-BISC^{22),23)}から全件取得し、それらのうちNDC新訂9版の分類記号(JAPAN-MARC 2009フォーマットのフィールド677)とNDLSHの件名標目(フィールド650:個人名件名標目, 658:一般件名標目)の両方をもつレコード934,251件を抽出した。フィールド650は、個人名からなる件名標目の場合に用いられ、それ以外の普通件名, 地名, 団体名, 家族名, および統一タイトルはすべてフィールド658に記録されている。

なお、それぞれのJ-BISC製作時点でのレコードであるため、その後に発生した件名標目の付け替えは反映されていない。以降の集計においてそれらは誤差となっているが、該当する件数は限られており、結果に大きな影響を及ぼすものではない。また、件名標目のIDはそのままで、件名見出し(ラベル)が変更されているものがある。こ

れらについては、IDを手がかりに最新の見出しに置換し、統一して用いた。

次に、抽出したNDLSHの件名標目に対して、Web NDL AuthoritiesのAPIを用いて、件名標目IDからそれぞれの件名標目の種別(普通件名, 地名などの種別)を判別した。この処理で該当データが見つからなかったものについては、人手で同システムを改めて検索し、件名の種別を調べた。その結果、普通件名78,346個(異なり数), 地名18,860個, 個人名25,155個, 団体名19,093個, 家族名1,010個, 統一タイトル2,794個であった²⁴⁾。また、各件名標目に割り当てられた代表分類記号については、Web NDL Authoritiesの一括ダウンロード用ファイルおよびAPIを用いて取得した。

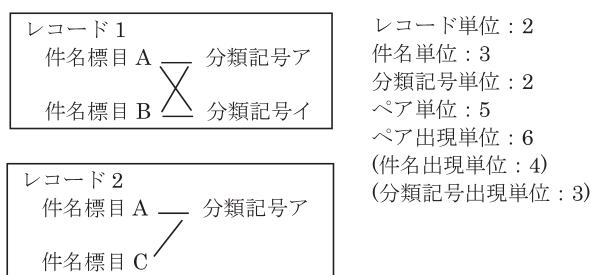
本研究では、得られた件名標目群のうち、普通件名と地名のみを取り上げる。NDLSHにおいて代表分類記号が示されているのは、例外的な統一タイトルなどを除いて普通件名と地名に限定されているからである。また、集計処理においては、件名標目IDは用いず、件名見出し(ただし、記号の正規化を加えたもの)の単位で実施した。同一件名見出しに対して、IDの付け替えなどが一部発生しているからである。

B. 普通件名の出現状況と代表分類記号の有無

対象とした書誌レコード群における普通件名の出現状況を表すには、複数の計測単位がある。第1図に示した例では、レコード1に件名標目2個と分類記号2個が付与されており、レコード2に件名標目2個と分類記号1個が付与されていることを表している。

この例示した事象においては、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」では2、件名標目の異なりの単位を指す「件名単位」では3(件名標目AからC)、分類記号の異なりの単位を指す「分類記号単位」では2(分類記号アとイ)と数えられる。同様に、件名標目と分類記号の組み合わせの異なりを指す「件名標目・分類記号ペア単位」では、総計5個のペア(件名標目Aと分類記号ア, Aとイ, Bとア, Bと

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



第 1 図 レコードにおける出現を表す各種の計測単位

第 1 表 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%)
細目なし	71 (0.1%)	256 (0.1%)	518 (0.1%)	
細目あり	63,930 (81.6%)	99,899 (49.3%)	334,975 (33.1%)	
件名当たりの平均		2.59 (SD: 5.89)	12.91 (SD: 69.47)	
最大数		315	5,971	
ペア当たりの平均			4.99 (SD: 50.38)	
最大数			5,647	

イ、C とア) となるが、ペア出現ごとの単位である「ペア出現単位」すなわちペア出現回数では 6 となる。後者は、個々のレコードに付与された件名標目数と分類記号数を掛け合わせた数をすべてのレコードで合計した値となる。また、件名標目の総出現回数（レコードへの総付与回数）を指す「件名出現単位」は 4、分類記号の総出現回数を指す「分類記号出現単位」は 3 と数えられる。したがって、レコード単位においては、単一レコードに複数の件名標目が付与されていても、それらをまとめてレコードの単位で集計され、件名標目の単位である件名単位では、件名標目がどれだけの異なる分類記号とペアを形成するのかは捨象さ

れ、集計値には反映されない。同じく、ペア単位ではペア異なり数のみ集計され、出現回数は捨象される。こうした複数の計測単位を適用して出現状況を多角的に把握することが必要である。

普通件名の出現状況を第 1 表にまとめた。各列は左から、件名単位、件名・分類記号ペア単位、ペア出現単位、そしてレコード単位を示している。

第 1 表に示した通り、普通件名が出現したのは総計 740,196 レコードであり、件名標目と分類記号とのペア総出現数（ペア出現単位）は 1,011,135 であった。件名標目の異なり数は 78,346、分類記号とのペア異なり数は 202,802 と

なった。これにより、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ことになる。また、件名標目当たり平均 2.59 個 (SD (標準偏差) 5.89), 最大で 315 個の異なる分類記号と組み合わせられている (第 1 表の下段)。同様に、件名標目当たりの平均ペア出現回数は 12.91 回 (SD 69.47), 最大で 5,971 回出現している。件名・分類記号のペア当たりでは平均 4.99 回 (SD 50.38), 最大 5,647 回の出現があった。

先に述べた通り、ペア出現単位の値は件名標目の総出現回数 (件名出現単位) とは若干異なる。レコードに分類記号が複数付与されているときには、ペア出現単位では分類記号の数だけ掛け合わせた値となるからである。しかしながら、今回の対象レコード群において分類記号が 2 個付与されたレコードは 12 にとどまり、それら分類記号と当該レコードに付与された件名標目との組み合わせによって、ペア出現単位の値は件名出現単位 (1,011,112) よりも 23 多い値となっている。この程度の差異であるため、両者を区別して扱う意味は殆どないと考える。

第 1 表には、件名標目における細目の有無、また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。次章において書誌レコードの分類記号と代表分類記号の一致を照合する範囲を、全体の中で確認しておくためである。

件名単位で見た場合、その 18.4% は細目をもたない、すなわち主標目のみからなる件名標目 (例: 「資料組織化」) 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。件名の異なり数で見れば、8 割は細目付き件名という状況である。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。さらに細目の有無を見ると、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。これは NDLSH の方針とも考えられ、基本的には細目付き件名標目には代表分類記号を割り当てていない。ペア出現単位についても上述の細目の有無の状況がそのまま反映してか、出現したペア (延べ出現回数) の 66.8% が代表分類記号をもち、その殆どは細目のない件名標目であった。ペア出現数の残り 33.2% は代表分類記号をもたず、その殆どは細目付き件名であった。

一方、細目があり、なおかつ代表分類記号をもつ件名標目が 22 個存在した。恐らくは例示的に代表分類記号を示したと推測される。それら 22 個の件名標目は、細目が付加されても主標目と同じ代表分類記号とされているものと、主標目とは異なる代表分類記号が割り当てられているものとの両方がある。「議会 || 補助機関」、「議会 || 委員会」は、「議会」と同じ代表分類記号「314」をもつ。それに対して、「聖書」の代表分類記号は「193」、細目が付いた「聖書 || 旧約」、「聖書 || 新約」はそれぞれ「193.1」、「193.5」となり、さらに「聖書 || 旧約 || 外典」「聖書 || 新約 || 外典」はいずれも「193.9」となり、主標目の代表分類記号の下位項目とされている。他方、「近代建築」(代表分類記号「523.05」) に対して、「近代建築 || 19 世紀」(同「523.06」)、「近代建築 || 20 世紀」(同「523.07」) のように、単純な分類記号の上位下位関係ではない例もある。

ペア出現回数の多い件名標目の上位 10 個と、出現ペア異なり数の多い件名標目 10 個を、第 2 表の左列と右列に、それぞれのペア出現回数、ペア異なり数、代表分類記号とともに示した。件名標目「人生訓」は 5,971 回ペアとして出現し、19 個の異なる分類記号と組み合わせられている。一方、「学位論文 || 抄録」は 5,647 回出現しているが、単一の分類記号「377.5」とのみ組み合わせられており、また当該件名標目に代表分類記号はない。ペア出現回数が多い件名標目においてもペア異なり数は相当程度に幅がある点が確認できる。

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

ペア異なり数最上位は「東日本大震災（2011）」であり、総計315の異なる分類記号と組み合わせられている。つまり、NDCの分類体系からみたときには多様な観点や文脈でこの「東日本大震災」が扱われていることを示しているといえよう。それに続くのが、「環境問題」、「太平洋戦争（1941-1945）」、「生死」、「コミュニケーション」であり、同様に多様な観点で取り上げられるトピックが並んでいる。「環境問題」と「太平洋戦争（1941-1945）」は左列にペア出現回数の多い件名標目としても出現している。なお、普通件名（代表分類記号をもたないものを含む）のペア出現回数とペア異なり数との相関係数は0.6910 ($p < .01$) であり、ペア出現回数が増加すればペア異なり数も増加するという相関が見て取れる。出現回数が突出して多い「人生訓」と「学位論文Ⅱ抄録」（これらはペア異なり数が少ない）を外れ値として除外し相関係数を求めると、その値は0.7591 ($p < .01$) まで上昇する。ペア異なり数が大きいほど、代表分類記号と一致しないペア数（異なり数）は増えることになり、最終的にはペア単位やペア出現単位での一致の程度に影響する。

また、代表分類記号をもたない件名標目のうち、細目をもたない主標目のみのものが71個ある（第1表）。「外資導入（アジア）」や「外資導入（イギリス）」、あるいは「磁気記録装置」、「劣化（紙）」などが該当する。これらが代表分類記号をもたない理由は不明である。

また、NDLSHにおいて代表分類記号が割り当てられているが、書誌レコード群において使用実績のない件名標目がある。これらは主に、書誌レコードには細目付き件名の形でのみ出現するが、その主標目のみによる件名標目としてNDLSHに登録されている。

続いて、先の書誌レコード群に出現した普通件名に対して、試みに細目を除去した主標目をもって集計を行った。もともと細目をもたないものはそのまま照合し、細目をもつものはそれを除去し、すべて主標目のレベルで統合した集計とした。これにより、細目付き件名のままでは除外される多数の事例を包含させて代表分類記号と照合できることになる。ただし、前述の通り、細目が付加されたときに主標目の代表分類記号とは異なる代表分類記号が割り当てられている事例があり、ここでの処理はそれらについて適切ではない

第2表 ペア出現回数、ペア異なり数の多い普通件名（上位10個）

ペア出現回数の多い件名	ペア出現回数	ペア異なり数	代表分類記号	ペア異なり数の多い件名	ペア出現回数	ペア異なり数	代表分類記号
人生訓	5,971	19	159	東日本大震災（2011）	1,301	315	369.31; 453.21
学位論文Ⅱ抄録	5,647	1	—	環境問題	2,560	314	519
住居址	3,028	20	202.5; 210.025	太平洋戦争（1941-1945）	2,953	212	210.75
太平洋戦争（1941-1945）	2,953	212	210.75	生死	1,057	169	114.2; 461.1
看護学	2,923	68	492.9	コミュニケーション	846	160	007.1; 361.45
環境問題	2,560	314	519	国際文化交流	748	156	002; 319
成功法	2,547	14	159	インターネット	1,474	148	547.483; 694.5
料理	2,344	93	596	データ伝送	1,975	139	547.48; 694.5
家庭教育	2,269	87	379.9	蔵書目録	813	139	029
コンピュータ	2,175	66	548.2	脳	1,033	134	491.171; 491.371

集計となる。集計の結果、主標目レベルでの件名標目の異なり数は15,021個となり、細目を伴った場合の78,346個に比べ大幅に減少した(第3表)。また、代表分類記号をもつものが、14,345個から14,866個へと若干増加を見せる。増加分は、細目付きで代表分類記号をもたなかった件名標目で、なおかつ主標目のみでの使用実績がなかったものが該当する。なお、こうした統合処理にも関わらず、主標目のみでも代表分類記号がないものが155残される結果となった。

全体的には主標目のレベルでは、代表分類記号をもつものが件名単位、件名・分類記号のペア単位、そしてペア出現単位においても、いずれも99%以上となった。ペア出現回数上位の件名標目は、「技術援助(日本)」(9,705回、169異なりペア)、「人生訓」(9,453回、42ペア)、「学位論文」(5,654回、1ペア)、「日本語」(5,605回、391ペ

ア)、「太平洋戦争(1941-1945)」(5,213回、276ペア)となり、第2表に含まれていないものもあった。「人生訓」は、主標目のみのペア出現回数5,971から、細目が付加されていたものを加えたペア出現回数9,453に増加し、分類記号とのペア異なり数も19から42に増加した。

C. 分類記号ベースの普通件名の出現状況

普通件名が出現したレコード群を対象に、分類記号の側から見たときの集計を行った(第4表)。あくまでも次章において代表件名標目との一致の程度を測ることに向けた集計であり、その範囲にとどまる。

出現した分類記号の異なり数は18,395個となり、そのうち33.7%に当たる6,208個の分類記号は、対応する代表件名標目をもっていた。件名標目ベースで見たときに、その20%弱のみ代表分

第3表 普通件名の出現状況：主標目レベル

	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	15,021	143,189	1,011,135	740,196
代表分類記号あり	14,866 (99.0%)	142,526 (99.5%)	1,009,048 (99.8%)	739,245 (99.9%)
代表分類記号なし	155 (1.0%)	663 (0.5%)	2,087 (0.2%)	951 (0.1%)
件名当たりの平均		9.53 (SD: 19.81)	67.31 (SD: 232.46)	
最大数		521	9,705	
ペア当たりの平均			7.06 (SD: 141.10)	
最大数			9,089	

第4表 分類記号ベースによる普通件名の出現状況

	分類記号単位	分類記号・件名ペア単位	ペア出現単位
総数	18,395	202,802	1,011,135
代表件名標目あり	6,208 (33.7%)	137,737 (67.9%)	770,875 (76.2%)
代表件名標目なし	12,187 (66.3%)	65,065 (32.1%)	240,260 (23.8%)
分類記号当たりの平均		11.02 (SD: 29.50)	54.97 (SD: 237.97)
最大数		1,046	9,316
ペア当たりの平均			4.99 (SD: 92.18)
最大数			5,647

NDSLH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

類記号をもっていたことに比べれば、この割合は大きいといえよう。代表分類記号をもつ普通件名の場合、総体としては、単一の件名標目に平均1.39個(SD 0.63)の代表分類記号が割り当てられているのに対して、単一の分類記号に平均3.88個(SD 4.72)の代表件名標目が割り当てられていることが反映しているものと考えられる。分類記号・件名のペア単位で見たときには、さらに増えて67.9%が代表件名標目をもつ。分類記号当たりの平均ペア異なり数は11.02(SD 29.50)であり、分類記号ごとに多様な件名標目との組み合わせが生じていることが分かる。最大数のペア異なり数1,046をもつ分類記号は「518.8 都市計画」であった。ペア出現単位で見たときにも、76.2%が代表件名標目をもち、分類記号当たりの平均ペア出現数が54.97(SD 237.97)、ペア当たりの平均出現数が4.99(SD 92.18)となっている。

ペア出現単位に関して補足すると、その値は、2個以上の件名標目が付与されているレコードが499,526件あり、それらと同一レコードに付与された分類記号の数を掛け合わせた合計値となっている。他方、レコード総数740,196件における分類記号の総出現数(分類記号出現単位)は740,208であり、ペア出現回数とは開きがある。しかし、ここでは件名標目ベースの集計と整合させるため、ペア出現単位の値を採用する。なお、

書誌レコードにおける件名標目付与数の分布は、件名標目が1個付与されているレコード511,586件、2個付与189,556件、3個付与35,911件、4個付与3,043件、5個付与91件、6個付与9件というものであった。

第5表は、左列にペア出現回数上位の分類記号、右列に出現ペアの異なり数上位の分類記号を、それぞれのペア出現回数、ペア異なり数とともに示した。分類記号「333.804 経済協力・援助: 論文集」が9,316回ペアとして出現し、268個の異なる件名標目と組み合わせられていた。次が「159 人生訓. 教訓」で、総計9,250回、ペア異なり数155をもって出現している。一方、ペア異なり数最上位は「518.8 都市計画」であり、それに続くのが、「369.26 老人福祉」, 「916 日本文学: 手記. ルポルタージュ」となり、多様な観点で取り上げられるトピックが並んでいると考えられる。これらはペア出現回数の多い分類記号として左列にも挙げられている。

D. 地名の出現状況と代表分類記号の有無

地名件名についても、普通件名と同様、出現の状況を集計した(第6表)。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。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

第5表 ペア出現回数、ペア異なり数の多い分類記号(上位10個)

ペア出現回数 数の多い 分類記号	項目見出し	ペア 出現回数	ペア 異なり数	ペア異なり 数の多い 分類記号	項目見出し	ペア 出現回数	ペア 異なり数
333.804	経済協力・援助: 論文集	9,316	268	518.8	都市計画	4,265	1,046
159	人生訓. 教訓	9,250	155	369.26	老人福祉	6,168	883
916	日本文学: 手記. ルポルタージュ	8,330	702	916	日本文学: 手記. ルポルタージュ	8,330	702
547.483	データ通信網	6,890	173	369.31	震災. 火山災害	3,350	651
377.5	学位 (大学. 高等教育)	6,370	20	519.1	公害・環境行政	2,693	648
369.26	老人福祉	6,168	883	369.27	障害者福祉	2,685	546
518.8	都市計画	4,265	1,046	369.3	災害. 災害救助	2,617	520
336	経営管理	4,168	299	369.4	児童福祉	2,453	497
210.0254	考古学(日本史)	4,058	158	519.4	水質汚濁. 海洋汚染	3,397	471
159.4	経営訓	3,858	82	501.6	工業動力. エネルギ	2,520	435

第6表 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
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%)
細目なし	2,859 (15.2%)	25,543 (67.2%)	127,494 (75.4%)	
細目あり	14,877 (78.8%)	8,122 (21.4%)	23,828 (14.1%)	
件名当たりの平均		2.02 (SD: 3.17)	8.96 (SD: 40.21)	
最大数		100	1,936	
ペア当たりの平均			4.44 (SD: 31.71)	
最大数			1,640	
<u>主標目レベル</u>	4,642	24,980	168,985	141,722
代表分類記号あり	923 (19.9%)	9,983 (40.0%)	79,936 (47.3%)	70,597 (49.8%)
代表分類記号なし	3,719 (80.1%)	14,997 (60.0%)	89,049 (52.7%)	71,125 (50.2%)
件名当たりの平均		5.38 (SD: 25.11)	36.40 (SD: 512.28)	
最大数		1,474	33,800	
ペア当たりの平均			6.76 (SD: 78.42)	
最大数			2,041	

は168,982回の出現であり、ペア単位の出現と変わらない。分類記号出現単位も141,723であり、1件のレコードに2個分類記号が付与されているケースが1つ存在するのみであった。

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、代表分類記号がある地名は6.0%にすぎず、大多数は代表分類記号をもたない。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。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。これらを個別にいくつか見てみると、その殆どにはNDLCの代表分類記号が示されており、NDLCを優先するという国立国会図書館の方針の現れと受け取ることができる。他方、細目をもつ地名のうち、201個には代表分類記号が示されており、単なる例示以上の数となっている。

全体的には、代表分類記号をもたないものは、やはり細目付き地名が大多数を占める。地名と分類記号のペア単位で見ても、代表分類記号をもつのは11.5%にとどまり、もたないものが88.5%を占める。ペア出現単位でも代表分類記号をもつものは10.5%、残り89.5%はそれをもたない。以上

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

第7表 ペア出現回数、ペア異なり数の多い地名（上位10個）

ペア出現回数の多い件名	ペア出現回数	ペア異なり数	代表分類記号	ペア異なり数の多い件名	ペア出現回数	ペア異なり数	代表分類記号
日本 経済	1,936	80	210.3	日本 社会	1,171	100	
日本 政治	1,608	50		日本 経済	1,936	80	
日本 歴史 古代	1,434	69		日本 文化	537	77	
日本 伝記	1,331	61		日本 遺跡遺物	690	70	
日本 紀行	1,176	54		日本 歴史 古代	1,434	69	210.3
日本 社会	1,171	100		日本 伝記	1,331	61	
日本 法令	905	3		多摩川	167	57	291.365
日本 歴史	900	53		日本 紀行	1,176	54	
千葉県 遺跡遺物	791	7		江戸	447	54	291.361
福岡市 遺跡遺物	752	3		日本 歴史	900	53	
				町田市	216	53	291.365

のように、地名については代表分類記号をもつものがいずれの単位においても低く、普通件名の状況とは異なる様相を呈している。

件名標目当たりの平均ペア異なり数2.02 (SD 3.17)、平均ペア出現回数8.96 (SD 40.21) は、普通件名の場合よりも小さい。ペア異なり数の最大数は100で「日本 || 社会」が該当し、ペア出現回数の最大数は1,936回で「日本 || 経済」であった。また、ペア当たりの平均ペア出現回数は4.44 (SD 31.71)、最大値は1,640回（「日本 || 経済」と「332.107」のペア）である。ペア出現回数の多い地名とペア異なり数の多い地名について、上位10個を第7表の左列と右列に示した。最大数をもつ先の件名標目に続いて、「日本 || 政治」、「日本 || 歴史 || 古代」というように、主標目「日本」をもつ件名標目がペア出現回数上位を占める。ペア異なり数についても、同様である。また、代表分類記号の欄が殆ど空白であるように、ペア出現回数等が多くても代表分類記号が割り当てられていないものが目立つ。

併せて、細目をもつものはそれを除去し、主標目のレベルで集計を行った（第6表の下段）。集計の結果、件名の異なり数は4,642個となり、もともと細目をもたない状態で出現した3,782個よりも増加した。つまり、細目を伴わない主標目のみの形では出現していなかった地名が、それだけ存在したということである。ただし、代表分類記号をもつ地名数はそれほど変化しておらず（923

個）、全体に対する割合も19.9%にとどまった。ペア単位で見たときには、40.0%が代表分類記号をもち、ペア出現単位についても同様に47.3%が代表分類記号をもち、残りはそれをもたない。主標目レベルにまとめたときには、代表分類記号をもつものが大幅に増加するが、全体に占める割合は普通件名の場合（いずれも99%以上）には至らず、異なる特徴を見せる。主標目レベルにまとめた後のペア出現回数上位の地名は、「日本」（33,800回、1,474ペア）、「中国」（4,901回、394ペア）、「東京都」（2,655回、189ペア）、「アメリカ合衆国」（2,434回、280ペア）、「北海道」（1,436回、160ペア）であった。

なお、地名件名については、上記のような代表分類記号の有無状況にあるため、分類記号ベースの出現状況の集計は省略する。

IV. 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

A. 一致数と一致率の平均

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。完全一致とは、書誌レコードの分類記号と代表分類記号とが完全に一致することを指す。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。

この完全一致に加えて、本研究では、前方一致を導入する。前方一致とは、完全一致はしないが、書誌レコードの分類記号が代表分類記号の下位区分であったときを指す。たとえば、第8表に示した説明用の事例では、件名標目「メタデータ」の代表分類記号は「014」と「014.3」の2個があり、これらに合致する書誌レコードは完全一致となる。他方、書誌レコードに出現した分類記号が「014.32」であれば、前方一致とみなす。つまり、完全一致を含まないが、上述の形で一致となるときを前方一致とする。これは、書誌レコードには代表分類記号よりもさらに特定性の高い分類記号が付与されているということであり、完全一致に準じたものとして一致とみなすことが妥当と考えるからである。NDCの補助表を適用して代表分類記号を展開した分類記号が書誌レコードに付与されているとき、それらを一致とみなすことをその典型例とする。ただし、件名標目の側でも、書誌レコードに付与された分類記号（上記の例では014.32）に対応するような細目がある場合、そうした細目を付加した件名標目（たとえば仮に「メタデータ || 基準」）は、通常、代表分類記号をもたないグループに属することになる点は留意しなければならない。つまり、そうした齟齬ともいえる可能性を捨象して無条件に前方一致を一致として採用してよいのか、場合によっては検討が必要となろう。

第8表に示した例では、件名単位で見たとき「メタデータ」も「資料組織化」もそれぞれが出現した書誌レコード群において代表分類記号と合致する分類記号をもつものが少なくとも1個以上あり、それゆえこれら2個の件名標目については

完全一致と数える。たとえ不一致となる分類記号をもつ書誌レコードが他に存在したとしても、件名単位では完全一致と数えられる。件名・分類記号のペア単位で見たときには、例示した事例では、件名標目「メタデータ」には4個のペアがあり、完全一致が2ペア、前方一致が1ペア、不一致が1ペアとなる。同様に、「資料組織化」は、完全一致ペアが1、不一致ペアが1となる。また、ペア出現回数では、「メタデータ」は、完全一致が8、前方一致が1、不一致が1となる。

次に、一致の程度を表す指標として一致率とその平均値を導入する。個々の件名標目ごとの一致率を計算し、出現したすべての件名標目について平均を求める。平均は、算術平均と加重平均を適用する。単純な算術平均には現れない、個々の件名標目の出現回数の差を反映させた加重平均の値を併せて示すことが適切と考えるからである。

先の第8表の事例を取れば、件名単位では、「メタデータ」の完全一致の一致率 $8/10=0.8$ 、前方一致を合わせたときの一致率 $9/10=0.9$ となる（第9表）。「資料組織化」は完全一致の一致率が $40/50=0.8$ 、前方一致を合わせても一致率0.8となる。これら2個の件名標目が対象とする件名全体を構成するとしたときには、一致率の算術平均は、完全一致の場合、 $(0.8+0.8)/2=0.8$ 、前方一致を合わせたときは、 $(0.9+0.8)/2=0.85$ となる。なお、3番目の件名標目が存在し、それについては代表分類記号に合致する書誌レコードがない、つまりすべて不一致であったときには、当該件名標目の一致率は0であり、3個の件名標目の平均一致率は、完全一致の場合、 $(0.8+0.8+0)/3=0.5333$ 、前方一致を合わせたときは、 $(0.9+0.8+0)/3=0.5667$ となる。このように一致率0の件名標目も含めて本稿では平均を算出する。

上記の一致率の算術平均では、「メタデータ」も「資料組織化」も一致率は0.8で同じ重みをもつ。しかしながら、それぞれの出現回数には5倍の開きがあり、この点を反映させた値とするべく下記の加重平均を並行して求める。

第8表 照合結果と一致数の説明例

件名	分類記号	照合結果	ペア出現回数
メタデータ	014	完全一致	2
メタデータ	014.3	完全一致	6
メタデータ	014.32	前方一致	1
メタデータ	015.2	不一致	1
資料組織化	014	完全一致	40
資料組織化	015	不一致	10

NDSLH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

第9表 平均一致率の計算例（件名単位とペア単位）

件名単位	「メタデータ」	「資料組織化」	
	一致率	一致率	算術平均
完全一致	8/10=0.8	40/50=0.8	(0.8+0.8)/2=0.8
完全一致+前方一致	9/10=0.9	40/50=0.8	(0.9+0.8)/2=0.85
	一致率×一致数	一致率×一致数	加重平均
完全一致	(8/10)×8=6.4	(40/50)×40=32	(6.4+32)/60=0.64
完全一致+前方一致	(9/10)×9=8.1	(40/50)×40=32	(8.1+32)/60=0.6683

ペア単位	「メタデータ」と「014」	「メタデータ」と「014.3」	「資料組織化」と「014」
	一致率		一致率
完全一致	2/10=0.2	6/10=0.6	40/50=0.8
完全一致+前方一致	3/10=0.3	7/10=0.7	40/50=0.8
	一致率×一致数		一致率×一致数
完全一致	(2/10)×2=0.4	(6/10)×6=3.6	(40/50)×40=32
完全一致+前方一致	(3/10)×3=0.9	(7/10)×7=4.9	(40/50)×40=32
			加重平均
			(0.4+3.6+32)/60=0.6
			(0.9+4.9+32)/60=0.63

$$\bar{x} = \sum_i x_i \cdot y_i / \sum_i y_i$$

x_i : 件名標目 i の一致率, y_i : 件名標目 i のペア出現回数

加重平均を求める際に分母を一致した回数とするのか、あるいは一致しない回数を含め出現したすべての回数とするのかという選択肢があり、後述のすべて不一致であった場合の処置とも一貫するよう、本研究では後者を採用する。上記の事例でいえば、完全一致の場合、分母を $8+40=48$ とするのか 60 とするのかという選択である。後者の 60 は全出現回数であり、これを分母とすることは、分子においても不一致であったものを実質的に計算に含めている、すなわち「メタデータ」について一致率 $0 \times$ ペア出現回数 $2=0$ 、「資料組織化」について一致率 $0 \times$ ペア出現回数 $10=0$ を含めていることになる。さらには、代表分類記号をもつ書誌レコードがない、つまりすべて不一致であった場合にも、当該件名標目の一致率は 0 とし、分母には当該件名標目のペア出現回数を加えて加重平均を求める。このような計算を採用し、不一致の数をも反映した平均値とする。なお、第9表では一致した部分のみ示しているため、「一

致率×ペア出現回数」ではなく、「一致率×一致数」としている。

それに対して、件名標目と分類記号のペア単位の一一致率とは、複数の代表分類記号をもつ場合にそれぞれを区別して計算するという処理に対応する。第8表の例では、件名標目「メタデータ」が代表分類記号「014」と「014.3」の2個をもつため、それぞれについて一致率を求める。ペア「メタデータ」と「014」の一一致率は、完全一致 $2/10=0.2$ 、ペア「メタデータ」と「014.3」は $6/10=0.6$ となる。ペア「資料組織化」と「014」の一一致率 $40/50=0.8$ を加え、これらペアの一一致率の算術平均を $(0.2+0.6+0.8)/6=0.2667$ と求める。分母を 6 とすることによって、一致率 0 のペアも計算に含めていることになる。

問題となるのはペア単位において前方一致を含めた場合である。「014.32」の1個が前方一致として加わるが、ペア「メタデータ」と「014」の一一致率、さらにペア「メタデータ」と「014.3」の一一致率において、この1個が重複して計算されてしまうことになる ($3/10$ と $7/10$)。これにより、一致率の平均は実際よりも多少なりとも大きな値となる。先の事例では、 $(0.3+0.7+0.8)/6=0.3$ となる。このような前方一致における重複算

出は、複数個の代表分類記号をもつ件名標目において、それら代表分類記号が上位項目とその下位項目という関係にある場合にのみ発生する。普通件名については42個の件名標目において、こうした関係が観察された。上記の「メタデータ」の事例以外に、「検査管理」が「509.6」と「509.66」、「トポロジー」が「415」と「415.7」をもつといった事例が見られた。逆にいえば、これら42個の件名標目を除いては、こうした重複算出は起こらない。

ペア単位の加重平均については、先の式で x_i をペア i の一致率、 y_i をペア i のペア出現回数とすることによって求める。前方一致を含めたときの重複算出の問題は、算術平均の場合と同じである。

ところで、和田らによる先行研究では件名単位とペア単位における一致率の算術平均のみ求めているが、示された値からはいずれについても一致率0、すなわち代表分類記号と一致する分類記号をもたない件名標目ないしペアは計算に含めていないと推測される。

B. 普通件名についての一致の程度

1. 代表分類記号との一致

先ず、細目付き件名標目をそのまま扱い、普通件名における一致数等を求めた。前述した通り22件名を除いては、細目付き件名標目は代表分類記号をもたないため、それらは照合対象に入らない。

第10表に得られた一致数を示し、第11表に平均一致率を示した。一致率およびその平均値は前節に記した式を用いて算出するが、すべて100倍した値にして示した（以降の表についても同じ）。

件名単位では、代表分類記号をもつ14,345件名のうち、完全一致したのは12,098件名（84.3%）、前方一致を加えたときは12,455件名（86.8%）が一致となった。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個以上もつことを表している。なお、これは件名異なり数を分母とする一致割合すなわち一致率ともいえる。残りの1,890件名（13.2%）については、代表分類記号と一致する分類記号が出現しなかった。件名・分類記号ペア単位では、13.7%が完全一致、前方一致を含

第10表 一致数：普通件名

	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		102,647		675,642		516,828	
完全一致数	12,098	(84.3%)	14,088	(13.7%)	383,277	(56.7%)	370,228	(71.6%)
完全一致数+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	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%)
完全一致数+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	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%)
完全一致数+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%)

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

第 11 表 平均一致率：普通件名

	件名単位	ペア単位	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
算術平均						
完全一致	54.94	7.68	76.46	30.31	53.79	5.61
完全一致+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
完全一致+前方一致	51.32	47.74	82.03	76.82	50.62	47.21

めても 20.7% が一致したにとどまった。つまり、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結果となった。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。さらにレコード単位で見ると、単一レコードに複数付与された件名のうち少なくとも 1 個が完全一致となるレコードが全体の 71.6%、前方一致を加えると 82.9% に達する。件名単位での完全一致・前方一致する割合の高さは当然とはいえ、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、代表分類記号にとって意味あるもの、すなわち 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となる。

これらについて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 54.94、前方一致を加えたとき 60.25 となった（第 11 表）。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、件名当たりの平均では 5 割以上の一致率、前方一致を含めると 6 割以上の一致率となる。それに対して、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。

次に、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。

第 12 表に、件名単位の一一致率（表内の「一致率（完全）」）に完全一致数を掛けた値が高い上位 10 件名を示した。件名「人生訓」は 5,971 回ペアとして出現し、5,578 回が完全一致であった。ペアの異なり数は 19、平均情報量は 0.4634 であった。ある件名標目についての平均情報量（エントロピー） H とは、下記の式で求められ、当該件名標目が複数の分類記号とペアを形成し、それぞれのペアにどのように出現回数が分布しているのか、つまり少数のペアに偏って出現しているのか、それとも多数のペアに均等に出現しているのかなど、偏りの程度を表す指標である。値が大きいほど多数のペアに均等に分散して出現しており、値が小さくなるほど少数のペアに偏って出現していることを表す。なお、式に含まれるペア i の出現確率 p_i は、「ペア i のペア出現回数 / 当該件名標目の総ペア出現回数」で求める。

$$H = -\sum_i p_i \log_2 p_i$$

p_i : ペア i の出現確率

件名標目「人生訓」は、ペア異なり数 19 と表に記載した件名群では少ない部類に属し、かつ比較的偏りのあるペア出現回数の分布を構成してい

第12表 件名単位の「一致率×完全一致数」値が高い上位10件名

件名	ペア 出現回数	ペア 異なり数	平均情報量	完全一致数	一致率 (完全)	完全+前方 一致数	一致率 (完全+前方)
人生訓	5,971	19	0.4634	5,578	93.42	5,956	99.75
株式相場	1,721	26	0.9562	1,669	96.98	1,669	96.98
成功法	2,547	14	0.8479	2,009	78.88	2,539	99.69
家庭教育	2,269	87	1.4905	1,854	81.71	2,052	90.44
料理	2,344	93	1.7966	1,754	74.83	2,197	93.73
健康法	1,567	61	1.0134	1,398	89.22	1,430	91.26
学習指導	2,173	56	1.6317	1,639	75.43	1,663	76.53
紀行	1,262	48	0.5695	1,197	94.85	1,198	94.93
人事管理	2,115	128	2.5046	1,529	72.29	1,557	73.62
マーケティング	1,703	111	1.8753	1,362	79.98	1,428	83.85

るため、平均情報量は小さい。なお、全件名標目の中で平均情報量が最大であるのは「蔵書目録」の6.0380（ペア出現回数813、ペア異なり数139）であり、「危機管理」の5.7980（276回、105ペア）、「記事索引」の5.7261（1,542回、121ペア）、「絵図」の5.7124（132回、66ペア）がそれに続く。一方、平均情報量が最小値0となる件名標目は多数あり、単一のペアとしてのみ出現するものがこれに該当する。

件名標目「人生訓」に戻ると、代表分類記号との完全一致数5,578、その一致率93.42、前方一致を含めた一致数5,956、一致率99.75であった。「一致率（完全）×完全一致数」の値が大きな順では「株式相場」、「成功法」、「家庭教育」と続く。「人生訓」、「成功法」、「家庭教育」、「料理」は、前掲の第2表にペア出現回数の多い件名標目として掲載されていたが、それら以外は新たな件名標目である。一致率（完全）が最高値100となる件名標目も多く（2,604個）、「鋼船」（ペア出現回数171）、「ウクレレ」（104回）、「近視」（99回）、「テニス」（97回）、「建築力学」（83回）などが該当する（いずれもペア異なり数1）。

第10表には、ペア出現単位とレコード単位の一一致数を併せて示してある。ペア出現単位では完全一致が全体の56.7%、前方一致を加えて66.7%が一致となり、出現した回数の半数強は代表分類記号と一致するケースであることになる。この割合は件名単位の一一致数の場合とペア単位の一一致数の場合との中間的な値となる。同様に、レコード

単位で見たときは完全一致が71.6%、前方一致を含めると82.9%が一致している。レコード当たり1.37個の普通件名が付与されていることに鑑みれば、レコードの7～8割において一致となることは意味あることと考えられる。

前方一致に関しては、当該レコードに付与された分類記号（それは件名標目に対する代表分類記号の下位分類項目となる）に対して、対応する件名標目が他に存在するかを併せて調べた。その結果、こうしたケースに該当するペアが総計3,546ペア存在した。たとえば、レコードに件名標目「簿記」（代表分類記号336.91）と分類記号「336.918」または「336.919」とがペアを形成する事例があるが、NDLSHにおいて「336.918」には代表件名標目「銀行簿記」、「工業簿記」などが、同じく「336.919」には「帳票」が割り当てられている。つまり、件名標目が表す特定性と分類記号が表す特定性が不整合であり、こうした事態を一致と見なしてよいかという問題である。この点については、「銀行簿記」、「工業簿記」などの代表件名標目では表しえないが「簿記」には属する、そうした事項が分類記号「336.918 各種簿記」には包含されると考え、特に問題視する必要はないものと本稿では考えた。

ところで、和田らによる報告では、完全一致の場合、一致率の平均（算術平均）は件名単位で57.14、ペア単位で40.31と示されている²⁾[p. 63]。これらの値が本稿での結果と異なるのは、彼らの研究ではa) 件名の種別を区分しておら

ず、すべての件名を一括して計算している、なおかつ対象期間が本研究と少し異なる、b) 出現回数25回以下の件名標目は一致率の計算から除外している、c) 一致率の平均を求める際に一致率0であった件名標目なりペアなりを含めていないという相違に帰因する。本研究での普通件名データについて、彼らと同様な計算で一致率0のものを除外して平均を求めたときには(ただし、他の条件は本研究の通り)、件名単位で57.91、ペア単位で44.61となり、ほぼ同等の結果となった。

彼らがb)の処置を採用したのは、“少数の書誌レコードによる影響が強く、本研究ではノイズであると判断し結果から除外した”²⁾[p. 62]と理由を記している。本研究では、こうした側面があることを配慮した結果、算術平均とは別に加重平均を採用している。試みに、上記b)の条件を採用して本研究の普通件名データについて一致率の平均を求めてみたところ(ただし、他の条件は本研究の通り)、件名単位の完全一致では算術平均56.52、加重平均39.85、ペア単位の完全一致では算術平均3.41、加重平均36.88となり、第11表に示した、b)を適用しないケースと比べ大きな変化は見られなかった。これゆえ、本研究では出現回数の閾値を設定し適用するという条件は採用しないこととした。

2. レコード内先頭出現の件名標目に限定した一致

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。『国立国会図書館件名作業指針』によれば、“【件名標目の付与順序】複数の件名標目を付与する場合、付与の順序は主な主題、副次的な主題の順とする。どれが主な主題か判断しがたい場合は、資料中で扱われている主題の順序に従う。”²⁵⁾[p. 3]とされている。この方針は、フィールド658に記録される普通件名、地名、団体名、家族名、および統一タイトルについて、件名の種別を問わず、当該資料の主な主題に対応するものから順に付与されるべきことを指示している。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。つまり、レコード内先頭の件名標目が普通件名でないものは採用しない、また普通件名であっても、その先頭出現のもののみ採用し、2番目以降は採用しないという方式で限定した。なお、分類記号については、複数付与されるケース自体が稀であるため、複数付与されている場合にもそれらすべてを採用し、先頭に限定することはしていない。

以上の限定を適用したときの一致数を、第10表の中段部分に示した。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。

これらの結果から、主な主題を表現している件名標目に限定したときには、不一致となる事例が減少し、代表分類記号に一致するものが多くを占めることが分かる。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(前方一致を含む)、ペア単位では30.31(完全一致)、31.79(前方一致を含む)となった(第12表の中列)。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。

3. 主標目に限定した一致

前掲の第3表に示した通り、本研究では、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ている。その結果、主標目に限定すれば、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代表分類記号と照合することができる。ただし、前述した通り、細目が付加されたときには主標目とは幾分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。それゆえ、主標目による照合結果は、あくまでも荒い近似と捉えるべきである。

照合の結果、細目付きの件名標目をそのまま扱ったときに比べて、521 件名 (=14,866-14,345) が新たに代表分類記号をもつことになったが、一致数は微増にとどまった (第 10 表の下段)。ペア単位では完全一致数、そして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 (第 11 表の右列)。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。この点より、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代表分類記号とする方策は適切ではないといえよう。

C. 分類記号ベースの普通件名についての一致の程度

分類記号をベースに、NDLSH が示すその代表件名標目と実際に書誌レコードに付与された件名標目との一致を照合した。その際、件名標目ベースの集計における「前方一致」に相当する一致を、分類記号ベースの集計においても導入する。

第 13 表 分類記号ベースの照合結果と一致数の説明例

分類記号	件名	照合結果	ペア出現回数
014	メタデータ	完全一致	2
014	資料組織化	完全一致	40
014.3	メタデータ	完全一致	6
014.32	メタデータ	継承一致	1
015	資料組織化	不一致	10
015.2	メタデータ	不一致	1

第 13 表に示した例は、先の第 8 表の事例を分類記号ベースに並べ替えたものである。分類記号「014」と件名標目「メタデータ」、「資料組織化」のそれぞれのペアは、代表件名標目と完全一致となる。同様に「014.3」と「メタデータ」のペアも完全一致である。それに対して、「014.32」は代表件名標目をもたず、そのままでは照合の対象外であるが、直上位桁の分類記号「014.3」が代表件名標目「メタデータ」をもつため、それと出現した件名標目が合致しており、上位桁からの継承による一致 (略して「継承一致」と見なすことができる。つまり、当該分類記号の上位桁に対応づけられた代表件名標目 (複数あるときにはそのいずれか) との合致である。なお、直近の上位桁に対応づけられた件名標目のみ照合対象に加えることとし、そのさらに上位の桁に割り当てられた件名標目を照合に加えることはしていない。たとえば、「014.3」に対応づけられた代表件名標目「メタデータ」と合致しないときに、さらに上位桁「014」に対応づけられた代表件名標目との照合を行うことはしていない。

第 13 表の次の事例である「015」には、代表件名標目として「エクステンションサービス (図書館)」、「図書館活動」など計 6 個が対応づけられているが、書誌レコードに付与された「資料組織化」は含まれていないため、照合結果は不一致となる。同様に「015.2」は代表件名標目として「参考図書」、「文献調査」、「レファレンスワーク」の 3 個があるが、「メタデータ」は含まれておらず、不一致となる。

上記のような継承一致ではなく、代表件名標目をもたない分類記号と組み合わせられた件名標目と、上位桁の分類記号に対応づけられている代表件名標目とが NDLSH において上位下位関係にあったときに一致と見なす方式も考えられる。たとえば、「014.32」には代表件名標目がなく、上位桁「014.3」がもつ代表件名標目と「014.32」と組み合わせられた件名標目とが一致ではなく、NDLSH における上位下位関係にあったときにも一致と見なすという方式である。しかしながら、NDLSH では上位下位関係が件名標目間に実際に

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

指示されている範囲は十分とはいえず¹⁰⁾、こうした方式での照合は限度があり、本稿では採用しない。

対象とした書誌レコード群において付与されている普通件名と代表件名標目とを分類記号ベースで照合してみると、分類記号単位では6,208個のうち、5,461(88.0%)が完全一致となった(第14表)。たとえ不一致となる件名標目をもつ書誌レコードが他に存在したとしても、分類記号単位では代表件名標目と合致する件名標目が少なくとも1個以上出現すれば、完全一致と数えられる。加えて、上記の継承一致となった分類記号が2,684個あり、完全一致と合わせると8,145の一致数となる。そのままでは代表件名標目をもたないとして数えられていた12,187個の分類記号から(第4表)、2,684個が継承一致ということで一致する側に新たに移行したことになる。そのため、第14表においては継承一致を先の完全一致に加えた数を角括弧で括っている。分類記号単位であるため、代表件名標目をもたない分類記号に対して1個でも継承一致となるペアが出現したときには、こうした移行が発生する。

分類記号・件名標目のペア単位では完全一致が10.2%(14,088)にとどまり、個々の分類記号に対して代表件名標目以外の多様な件名標目がベ

アを形成していることが見て取れる。継承一致を含めたときには21,263ペアの一致となり、7,174ペアという相当数の増加が見られる。ペア出現単位については、代表件名標目をもつものの約半数が完全一致であった。また、継承一致を含めた場合にも、大幅に一致件数が増加(67,321増加)している。やはり、代表件名標目と合致する件名標目の付与される場合が相当程度を占めるといえよう。

レコード内先頭出現の件名標目に限定したデータを用いて、分類記号ベースで集計したときにも、概ね類似の傾向が見られた(第14表の下端)。ここでは、あくまでもレコード内先頭出現の普通件名と分類記号との組み合わせという限定化であり、付与された分類記号の側からの限定とはしていない。照合の結果、分類記号単位、ペア単位、ペア出現単位のいずれにおいても、限定しないときに比べ完全一致となるものの割合が増加している。加えて、継承一致も相当数発生している(分類記号単位2,246、ペア単位4,664、ペア出現単位51,540の一致数増加)。これらにより、レコード内先頭出現の件名標目に限定することは、分類記号ベースの集計においても一致する割合の上昇に有効であることが示された。

次に、分類記号ベースにおける一致率の平均

第14表 分類記号ベースの一致数：普通件名

	分類記号単位		分類記号・件名ペア単位		ペア出現単位
代表件名標目あり	6,208 [8,892]		137,737 [144,912]		770,875 [838,196]
完全一致数	5,461 (88.0%)		14,088 (10.2%)		383,277 (49.7%)
完全一致数+継承一致数	[8,145]		[21,263]		[450,598]
不一致数	747 (12.0%)		123,649 (89.8%)		387,598 (50.3%)
レコード内先頭出現	5,922 [8,168]		57,295 [61,959]		538,294 [589,834]
完全一致数	5,368 (90.6%)		13,386 (23.4%)		352,664 (65.5%)
完全一致数+継承一致数	[7,614]		[18,050]		[404,204]
不一致数	554 (9.4%)		43,909 (76.6%)		185,630 (34.5%)

注：[]内の値は、継承一致を完全一致に加えた数

第15表 分類記号ベースの平均一致率：普通件名

	分類記号単位	ペア単位	レコード内先頭出現件名	
			分類記号単位	ペア単位
算術平均				
完全一致	33.89	2.08	45.69	6.02
完全一致+継承一致	67.23	2.58	74.88	7.35
加重平均				
完全一致	27.88	16.29	46.18	27.97
完全一致+継承一致	35.31	41.70	55.24	64.49

を、件名標目ベースのときと同じ方式で計算した(第15表)。分類記号単位の場合、算術平均では33.89(完全一致)、67.23(継承一致を含む)、加重平均では27.88(完全一致)、35.31(継承一致を含む)となった。一方、ペア単位では、算術平均で2.08(完全一致)、2.58(継承一致を含む)と極めて低い値であったが、加重平均にして16.29(完全一致)、41.70(継承一致を含む)となり、やや持ち直した値となった。全般的には、件名標目をベースにした集計結果と似た傾向といえよう。ただし、総じて値は低く、分類記号をベースに集計したときには多様な件名標目とその組み合わせとして出現し、代表件名標目と一致する事例に限られている状況が確認できよう。なお、和田らによる報告では、完全一致の場合、一致率の平均(算術平均)は分類記号単位で61.79、ペア単位で9.01とされている²⁾[p.63]。この計測の条件については前節で述べた通りであるが、本研究での算出結果と大きく異なる。恐らくは、複数の要因が組み合わさった結果と推測される。

併せて、レコード内先頭出現の件名標目に限定した場合について、平均一致率を求めた(第15表の右列)。分類記号単位、ペア単位のいずれの場合も、先の限定しないときに比べ、相当程度の上昇が認められた。しかしながら、同じくレコード内先頭出現の件名標目に限定した上での件名標目をベースにした集計結果の値には及ばなかった。

D. 地名件名についての一致の程度

地名による件名標目について、その一致数を

第16表に、一致率の平均を第17表に示した。代表分類記号をもつ1,124件名のうち、54.5%(613件名)が完全一致となり、前方一致を含めると55.5%が一致となった。件名・分類記号のペア単位では、14.4%が完全一致、前方一致を含めると15.8%の一致であった。同様に、ペア出現単位で見るときには、完全一致が42.0%、前方一致を含めると43.4%であった。これらを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、代表分類記号をもつ件名標目数、ペア異なり数、ペア出現回数いずれも総数としては減少している(第16表の中段)。それに対して、完全一致、前方一致を含めた一致の占める割合は、件名単位、ペア単位、ペア出現単位のいずれにおいても、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するが、ペア異なり数とペア出現回数は大きく上昇する(第16表の下端)。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、ペア異なり数とペア出現回数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(「地名のものと細目」)が付いたときに

NDSLH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

第16表 一致数：地名件名

	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	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%)
完全一致数+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	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%)
完全一致数+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	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%)
完全一致数+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%)

第17表 平均一致率：地名件名

	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	32.45	8.38	38.65	12.79	20.76	1.92
完全一致+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
完全一致+前方一致	29.56	29.37	37.59	37.38	3.09	2.89

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。地名による件名については、主標目にまとめることによる一致の増加は期待しえない。

一致率の平均を求めると、件名単位の算術平均では完全一致が32.45、前方一致を加えたときが33.61、ペア単位では完全一致が8.38、前方一致が8.52となった(第17表)。件名単位では普通件名に比べて低い値となったが、ペア単位については同程度の値であった。一方、加重平均では件名単位で完全一致が27.98、前方一致を含めて29.56、ペア単位の完全一致が27.88、前方一致が

29.37となり、普通件名の場合と比べ総じて低い値となった。

それに対して、レコード内先頭出現に限定したときには、一致率の平均は上昇を見せる。ペア単位の算術平均の値を除いて、35～40の値となった。しかしながら、普通件名の場合のような大幅な上昇は見られなかった。元々、レコード当たりの地名件名の付与数は少なく(平均1.19)、複数の地名件名を付与しているケースが限られているため、レコード内先頭出現に限定しても大きな変更とはならない事態を反映している。併せて、主標目のレベルにまとめたときの平均一致率を求め

たが、この処置を適用しない場合よりも、大幅に低い値となった。

V. 考 察

A. 相関係数による出現状況の考察

書誌レコードに付与された分類記号と代表分類記号との一致数および一致率を前章で求めたが、それらの値とペア出現回数やペア異なり数との間には相関が見られるだろうか。仮に高い相関が見られるならば、その規則性を活用することができる。たとえば、不一致ペアや代表分類記号をもたないペアからの有効なペアの抽出などへの適用が考えられる（次節を参照）。併せて、項目間のすべての相関係数を求めることで、それぞれの項目値の意味と関係を再確認することができる。

件名単位による検討を先ず試みるべく、個々の件名標目が形成するペアの総出現回数（ペア出現回数）、分類記号とのペアの異なり数、各ペアの出現分布を表す平均情報量と、一致数や一致率との相関を検証した。ここでは、代表分類記号をもつ14,345の普通件名について先の項目間の相関係数を求めた（第18表）。

件名標目のペア出現回数とペア異なり数とは相関係数0.704とかなりの相関が見られ、ペア出現回数が増えれば多様な分類記号とのペアも生成されうる状況が見て取れる。ペア異なり数が増えれば、多数のペアに出現回数が分散し平均情報量も増加するという状況も併せて示されている（相関係数0.629）。

完全一致数そして前方一致を加えた一致数

は、ペア出現回数と高い相関（相関係数0.881と0.938）をもち、かつペア異なり数ともある程度の相関を見せている（0.508と0.534）。すなわち、ペア出現回数やペア異なり数が増えれば、一致数も増加するという傾向がある。ただし、平均情報量とは殆ど相関がない。他方、完全一致の一致率と前方一致を加えた一致率は、ペア出現回数、ペア異なり数とは相関がなく、平均情報量とは弱い負の相関（-0.399と-0.334）を見せている。平均情報量が小さい、すなわち偏った分布を示す方が一致率が高いという状況を示している。また、一致率は一致数とは殆ど相関がない点も示されている。

なお、相関係数は外れ値の影響を受けやすい特性があるため、ペア出現回数が突出して大きい件名標目「人生訓」を外れ値として除外した上で項目間の相関係数を求めてみたが、値が若干変化するとはいえ、大きな変動は見られなかった。レコード内先頭出現の件名標目に限定したときも、項目間の相関係数には大きな変化はなかった。

併せて、件名標目ごとの一致数を従属変数とし、ペア出現回数、ペア異なり数、平均情報量を独立変数とした重回帰分析を行った。その結果、先の相関係数の高さから推測できるように、高い決定係数が得られる。完全一致数の場合に、自由度調整済み決定係数0.801となり、前方一致を含めた場合には同0.913であった。ペア出現回数およびペア異なり数から、多重共線性が含まれているとはいえ、ある程度の確度をもって一致数は予測可能であるといえよう。

第18表 件名単位による各項目間の相関係数：普通件名

	ペア 出現回数	ペア 異なり数	平均 情報量	完全 一致数	完全+ 前方一致数	一致率 (完全)	一致率 (完全+前方)
ペア出現回数	—						
ペア異なり数	.704**	—					
平均情報量	.269**	.629**	—				
完全一致数	.881**	.508**	.140**	—			
完全+前方一致数	.938**	.534**	.164**	.922**	—		
一致率(完全)	.017*	-.124**	-.399**	.125**	.077**	—	
一致率(完全+前方)	.061**	-.076**	-.334**	.132**	.134**	.892**	—

** : $p < .01$; * : $p < .05$

NDSLH 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

次に、普通件名と分類記号のペア単位 102,647 に移り、個々のペアの出現回数（個別ペア出現回数）、当該件名標目が形成するペアの総出現回数（件名総ペア出現回数）、完全一致数、前方一致を加えた一致数、完全一致の一致率、前方一致を含めた一致率について、相互の相関係数を求めた（第 19 表）。完全一致数と個別ペア出現回数が相関係数 0.895、前方一致を含めた一致数と個別ペア出現回数が相関係数 0.961 という高い相関を示した。一方で、これら一致数と件名標目の総ペア出現回数とは相関がなかった。一致率についても、個別ペア出現回数、件名総ペア出現回数、さらには一致数とも弱い相関しか見られない。

一致数を従属変数とし、個別ペア出現回数、件名総ペア出現回数、一致数を独立変数とした重回帰分析を併せて行った。その結果、先の相関係数の高さから推測できるように、高い決定係数が得られる。完全一致数の場合は自由度調整済み決定係数 0.802、前方一致を含めた場合には同 0.923 となった。

B. 不一致ペアおよび代表分類記号をもたないペアからの有効なペアの抽出

代表分類記号をもつ件名標目においては、前方一致となるものを除き、代表分類記号と異なる分類記号はすべて不一致となる。しかしながら、こうした不一致とされたペアの中には有効なペア、すなわち代表分類記号として採用されていないが妥当または有用な分類記号を示したペアが含まれている可能性がある。そのような有効なペアを抽出し、代表分類記号への追加候補として検討する

こと、あるいは代表分類記号とは異なる性格をもつ有用なペアとして分類作業・件名標目付与作業の支援システム等に組み入れることなどが考えられる。そこで、有効なペアをどのようにして抽出するかが課題となる。

前節で行ったペア単位に関する回帰分析からは、ペアごとの一致数であれば当該ペアの出現回数（個別ペア出現回数）、さらには「個別ペア出現回数×ペア出現率」と高い相関を有することが判明している（後者との相関係数 0.971）。なお、ペア出現率は、「個別ペア出現回数 / 件名総ペア出現回数」で求める。これらゆえ、個別ペア出現回数と、ペア出現率すなわち該当する件名標目が形成するペア出現全体における当該ペアの占める割合を考慮した値を手がかりにすることが適切であると仮定し、この値の大きなペアを取り出した。上位 20 ペアを第 20 表に示し、表の左から件名標目、不一致となった分類記号と項目見出し、当該ペアの出現回数（個別ペア出現回数）、出現率とその件名標目自体の総ペア出現回数、それに NDSLH で割り当てられた代表分類記号とその項目見出し、そしてこの代表分類記号と実際に付与された分類記号との完全一致数および一致率を記載した。

表に記載されたペア群において顕著な点の 1 つは、分類記号「916 記録、手記、ルポルターージュ」の出現であり、計 5 個の件名標目（闘病看病、太平洋戦争、引揚者、日中戦争、シベリア抑留）において登場している。分類記号「916」は対象となるトピック自体を表す項目ではなく、文学さらにはその形式としての側面を表す項目であり、複

第 19 表 ペア単位による各項目間の相関係数：普通件名

	個別ペア 出現回数	件名総ペア 出現回数	完全一致数	完全+ 前方一致数	一致率 (完全)	一致率 (完全+前方)
個別ペア出現回数	—					
件名総ペア出現回数	.116**	—				
完全一致数	.895**	.074**	—			
完全+前方一致数	.961**	.095**	.931**	—		
一致率 (完全)	.221**	-.144**	.272**	.248**	—	
一致率 (完全+前方)	.233**	-.152**	.264**	.262**	.971**	—

**：p<.01

数のトピックに適用されていることが分かる。このように、分類記号「916」はこれら件名標目の代表分類記号とはなりえない。同様に、件名標目「哲学」、「芸術」、「科学」に対して、それぞれ「104」、「704」、「404」という論文集・評論集・講演集を表す分類記号との組み合わせが目立つ。これらもトピックの取り上げ方すなわち形式の側面を含んだ分類記号であり、先の件名標目の代表分類記号として示すには適さない。むしろ、「哲学」論文集」など、細目が付加されたときの代表分類記号となる。

これに対して、件名標目「通信網」と「547.483 データ通信網」、「日本人」と「361.42 地方性. 国民性. 民族性」、「馬」と「788.5 競馬」、「電子出版」と「021.49 著作. 編集: コンピュータによる編集」、「性風俗」と「367.9 性問題. 性教育」などは、それぞれの件名標目の代表分類記号として追加することが適切と思われる。

残りは判断に迷う事例である。「サラリーマン」と「159.4 経営訓」のペア、「恋愛」と「159.6 女性のための人生訓」のペアは、それぞれ件名標目と分類記号（分類項目）とが対象者とそれに対する訓示、あるいはその主たるトピックという関係にある。つまり、件名標目と分類記号との概念的な同等性はないが、密接な関係をもつことも事実である。こうした関係にあるものを、代表分類記号に加えることが適切かは別途の検討を要する。前者のペアは、「人生訓」サラリーマン」とその代表分類記号「159.4」とすべきかもしれない。「老人性認知症」と「369.26 老人福祉」とのペアについても同様に、微妙な対応関係にある。

ここでは上位 20 ペアのみ取り上げたが、ペアをなす件名標目と分類記号の間に大きく齟齬がある事例は含まれていない。よって、20 位以降のペアも同様に調べる価値があると考えられるが、同時に有効な候補ペアの抽出方法の洗練化が検討されるべきであろう。

次に、代表分類記号をもたない普通件名に対して、上記の不一致ペアの場合と同様、「個別ペア出現回数 × ペア出現率」の値を手がかりに有効なペアの抽出を試みる。数の上で多数を占める代

表分類記号をもたないものに関して、その中から有効なペアすなわち代表分類記号の候補となるものを抽出するという課題に対する一つの単純な試みである。先の値による上位 20 ペアを第 21 表に示すが、すべて細目をもつ件名標目であった。

それらを個別に見てみると、件名標目と分類記号のペアは、殆どが概念的な同等性を表す組み合わせと判断される。

ただし、「技術援助（日本）」インドネシア」、「技術援助（日本）」中国」など、主標目「技術援助（日本）」と地名細目からなる件名標目については、主標目に対応する分類記号「333.8」に形式区分「04」が付いた分類記号が組み合わせられており、多少ともずれがある。逆にこれらの件名標目の代表分類記号であれば、地理区分を伴った分類記号が適切となる。また、「学位論文」抄録」、「科学教育」日本」、「法人税」日本」については、それぞれの主標目に該当する分類記号が組み合わせられている。地理区分を付加した分類記号とする方針がないときには、これで適切な分類記号となる。一方、「経済政策」日本」には、「332.1 日本経済史・事情」ではなく、時代を昭和時代後期・平成時代にさらに限定した「332.107」が組み合わせられている。このように代表分類記号とするには幾分修正が必要な組み合わせが見られるが、多くのものはそのまま代表分類記号として採用しうる適切な組み合わせと考えられる。従って、不一致のペアと同様、代表分類記号をもたない件名標目についても 20 位以降のペアも調べる価値があると考えられる。あるいは、レコード内先頭出現の件名標目に限定した上で、同様に上位ペアから調べることも有効かもしれない。

VI. おわりに

本研究では、国立国会図書館作成の書誌レコードに付与された NDLSh 件名標目と NDC 分類記号の組み合わせの中で、NDLSh が示す個々の件名標目に対応する代表分類記号が付与されている数および割合を調査し、その結果を分析した。このような調査には、和田らによる先行研究があるが、彼らの成果を踏まえつつも、さらに展開した

NDSLH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

第20表 不一致とされた普通件名と分類記号のペア（「個別ペア出現回数×ペア出現率」値が高いペア上位20）

件名	不一致の分類記号と 項目見出し	個別ペア 出現回数	ペア 出現率	件名総ペア 出現回数	代表分類記号と 項目見出し	完全 一致数	一致率 (完全)
闘病看病	916 記録. 手記. ル ポルターージュ	1,198	96.54	1,241	490.14 医学と心理	0	0.00
太平洋戦争 (1941-1945)	916 記録. 手記. ル ポルターージュ	1,584	53.64	2,953	210.75 太平洋戦争 (1941-1945)	655	22.18
サラリーマン	159.4 経営訓	678	83.60	811	361.84 中間階級: サ ラリーマン	60	7.40
通信網	547.483 データ通信網	1,001	53.44	1,873	547.2 通信方式. 通信 線路. 通信網; 694 電気通信事業	24	1.28
恋愛	159.6 女性のための人 生訓	973	51.16	1,902	141.62 愛情; 158 倫 理学. 道徳: その 他の特定主題; 384.7 性風俗	238	12.51
古墳	210.0254 考古学	452	68.28	662	210.32 大和時代: 氏 姓国家時代, 古墳 時代	96	14.50
文章	816 文章. 文体. 作 文 (日本語)	353	86.73	407	801.6 文章論. 文体 論. 修辭学	11	2.70
プログラミング (コンピュータ)	547.483 データ通信網	682	32.85	2,076	007.64 コンピュータ プログラミング	855	41.18
引揚者	916 記録. 手記. ル ポルターージュ	229	79.24	289	369.37 戦災者・被災 者・引揚者保護	46	15.92
日中戦争 (1937- 1945)	916 記録. 手記. ル ポルターージュ	322	53.05	607	210.74 日中戦 (1937-1945)	118	19.44
日本人	361.42 地方性. 国民 性. 民族性	301	55.54	542	281 日本の列伝; 382.1 日本の風俗・ 習慣, 民俗, 民族; 469.91 人種誌: 日 本	14	2.58
老人性認知症	369.26 老人福祉	407	40.90	995	493.758 老人精神病: 老人性痴呆	328	32.96
シベリア抑留	916 記録. 手記. ル ポルターージュ	186	85.71	217	210.75 太平洋戦 (1941-1945)	24	11.06
馬	788.5 競馬	229	68.77	333	489.8 奇蹄目: うま, さい, ばく; 645.2 家畜・畜産動 物各論: 馬	60	18.02
哲学	104 哲学: 論文集, 評論集, 講演集	274	54.80	500	100 哲学	116	23.20
芸術	704 芸術: 論文集, 評論集, 講演集	240	62.02	387	700 芸術. 美術	4	1.03
科学	404 自然科学: 論文 集, 評論集, 講演 集	240	61.86	388	400 自然科学	92	23.71
電子出版	021.49 著作. 編集: コンピュータによ る編集	212	63.66	333	023 出版	36	10.81
スマートフォン	007.64 コンピュータ プログラミング	254	52.48	484	547.62 移動無線: 694.6 電話	89	18.39
性風俗	367.9 性問題. 性教育	158	83.16	190	384.7 性風俗	17	8.95

第21表 代表分類記号をもたない普通件名と分類記号のペア（「個別ペア出現回数×ペア出現率」値が高いペア上位20）

件名	分類記号と項目見出し	個別ペア出現回数	ペア出現率	件名総ペア出現回数
学位論文 抄録	377.5 学位	5,647	100.00	5,647
人生訓 女性	159.6 女性のための人生訓	1,309	98.57	1,328
国勢調査 日本	358.1 人口統計, 国勢調査 (日本)	1,021	99.61	1,025
人生訓 サラリーマン	159.4 経営訓	1,006	98.53	1,021
英語 会話	837.8 英会話	905	86.85	1,042
真宗 感想説教	188.74 真宗 [浄土真宗]: 法話・語録, 説教集	712	99.03	719
成功法 サラリーマン	159.4 経営訓	694	98.30	706
プロテスタント教会 感想説教	198.34 プロテスタント, 新教: 信仰録, 説教集	687	91.36	752
憲法 日本	323.14 日本国憲法	861	69.72	1,235
禅宗 感想説教	188.84 禅宗: 臨済宗, 曹洞宗, 黄檗宗: 法話・語録, 説教集	563	98.08	574
技術援助 (日本) インドネシア	333.804 経済協力, 経済援助: 論文集, 評論集, 講演集	589	92.90	634
英語 研究指導	830.7 英語: 研究法, 指導法, 教育	569	91.19	624
日本語 研究指導	810.7 日本語: 研究法, 指導法, 教育	518	93.00	557
日本文学 歴史 明治以後	910.26 日本文学: 近代: 明治以後	571	83.24	686
科学教育 日本	375.4 科学教育	483	97.38	496
技術援助 (日本) 中国	333.804 経済協力, 経済援助: 論文集, 評論集, 講演集	474	92.22	514
法人税 日本	336.983 取得税: 所得税, 法人税	530	82.43	643
経済政策 日本	332.107 日本経済史・事情: 昭和時代後期, 平成時代	465	92.63	502
技術援助 (日本) フィリピン	333.804 経済協力, 経済援助: 論文集, 評論集, 講演集	431	95.57	451
金融 日本	338.21 日本金融史・事情, 銀行史・事情	483	81.31	594

詳細な集計と分析を実行した。具体的には、どのような範囲（条件）において付与された件名標目や分類記号が代表分類記号、代表件名標目と一致する割合が高いのか、範囲を変動させて集計と分析を行った。併せて、実態を多角的に把握するため、複数の集計単位を設定し、かつ一致の程度についても複数の指標を導入し表した。

代表分類記号の対応づけの状況が件名の種別によって大きく異なるため、本研究では普通件名と地名による件名に分けて調査し、下記のような結果を得た。

1) 普通件名に関する件名標目ベースの照合結果: a) 代表分類記号をもつ件名標目は全体の20%以下にとどまり、それらの殆どは細目をもたない。件名・分類記号のペア単位では、約半数が

代表分類記号をもち、それらと書誌レコードに付与された分類記号とを照合するのは可能である。b) これら代表分類記号をもつ件名標目について照合したところ、件名単位で80%以上、件名・分類記号ペア単位で10%強、ペア出現単位で60%弱が完全一致となった。分類記号の前方一致まで含めたときには、件名単位、ペア単位、ペア出現単位のいずれにおいても一致となるものが増加した。特にペア出現単位における一致となる割合の高さは、結果的にせよ、代表分類記号（またはその下位分類記号）を資料に付与する場合が相当程度あることを示している。c) 一致率の平均を求めたところ、件名単位では、たとえば完全一致の場合、算術平均で54.94（件名単位）、7.68（ペア単位）、加重平均で39.92（件名単位）、

37.08 (ペア単位) となった。なお、ここでは不一致のもの、一致率0のものを含めた算出法としている。d) 対象をレコード内先頭出現の件名標目に限定したとき、一致数および平均一致率の大幅な上昇が見られた。対象となる個々の資料の主な主題を表すとされる件名標目への限定は、代表分類記号(またはその下位分類記号)との一致を顕著化する。それに対して、e) 細目を除去し主標目のレベルですべての件名標目を照合したところ、これらの値はむしろ低下した。

2) 普通件名に対する分類記号ベースの照合結果: 上記と同じデータを分類記号の側から集計したときには、a) 代表件名標目をもつ分類記号は30%強であった。b) 書誌レコードに付与された件名標目との照合では、分類記号単位で90%弱、分類記号・件名ペア単位では10%強、ペア出現単位で50%弱が完全一致となった。c) 継承一致という独自の照合を導入すると、一致となるものが相当数増加した。d) 平均一致率を求めると、上記の件名標目ベースの場合にはとどかない値となった(分類記号単位で完全一致の算術平均33.89、加重平均27.88)。これは単一の分類記号に多数の件名標目の組み合わせ(平均ペア異なり数11.02個)が出現している状況を反映しており、代表件名標目と一致する範囲はかなり限定的であることを示している。

3) 地名件名に関する照合結果: a) NDC代表分類記号をもつ地名件名は僅か6%にとどまり、件名・分類記号のペア単位でも10%にとどまる。極めて限定的にNDC代表分類記号が示されている状況にある。b) これらのうち、完全一致となるのは、件名単位で50%強、ペア単位で15%程度、ペア出現単位で40%強であった。c) 平均一致率も、全般的に普通件名の場合に比べて低い値となった(件名単位における完全一致の算術平均32.45、加重平均27.98)。d) 細目を除去し主標目レベルで照合すると、一致数の占める割合は件名単位で横ばい、ペア単位で減少となり、平均一致率においてはいずれも大幅に低下する。地名の場合、付加される細目(「地名のもの細目」)に依存して適切な分類記号が決まるとい

特徴がうかがえた。

以上のことより、普通件名、さらにはレコード内先頭出現に限定したとき、代表分類記号との一致数および平均一致率は最も大きな値をとることが分かった。また、全体を通して、分類記号の前方一致・件名標目の継承一致の適用、レコード内先頭出現への限定化、細目を除去し主標目レベルへの統合化について、それぞれの効果の程度が確認できた。

なお、普通件名・地名件名以外の件名標目(個人名、団体名、家族名、統一タイトル)については、原則的に代表分類記号が付与されていないため、本研究では扱わなかった。また、代表分類記号としてはNDC分類記号のみを取り上げたが、NDLC分類記号について同様に集計や分析を行うことは可能である。

上記のように本研究で確認できた状況を踏まえて、今後は下記の事項の検討などが望まれよう。

1) 代表分類記号とは不一致となる件名標目と分類記号のペア、代表分類記号をもたない件名標目と分類記号のペアから、有効なペアをいかにして抽出するのか検討が必要である。本研究では、この点に関して単純な方式を試みたにとどまった。特に細目付き件名標目は現状では殆どが代表分類記号をもたないため、それらに対する代表分類記号の割り当ては重要な検討課題である。

2) 件名標目とその代表分類記号、分類記号とその代表件名標目という対応づけをいかに活用すべきか、有効な活用法の検討が望まれる。さらには、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を先の対応づけに加えた活用法、特にOPAC等の検索システムにおける活用法は検討に値しよう。主題表現と検索の仕組みとしての分類法と件名標目表、そして分類記号と件名標目とが、現状では十分に活かされていないと考えるからである。

注・引用文献

- 1) 嶋田真智恵. 国立国会図書館件名標目表(NDSLH)の改訂作業と今後について. 情報の科学と技術. 2007, vol. 57, no. 2, p. 73-78.

- 2) 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. 札幌, 2010-10-9/10, 日本図書館情報学会, 2010, p. 61-64.
- 3) 石田栄美. 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: 駿河台大学文化情報学部紀要. 2005, vol. 12, no. 1, p. 1-11.
- 4) Frank, Eibe; Paynter, Gordon W. Predicting Library of Congress Classifications from Library of Congress Subject Headings. *Journal of the American Society for Information Science and Technology*. 2004, vol. 55, no. 3, p. 214-227.
- 5) 村上幸二, 村上晴美. NDCカテゴリを用いた複数件名検索システムの開発. 情報学: *Journal of Informatics*. 2007, vol. 4, no. 2. <http://kiyo.info.gsucc.osaka-cu.ac.jp/JI/article/view/58/53>, (accessed 2015-05-20).
- 6) 村上晴美, 平田高志, 上田洋. Subject World: 主題の世界. 情報の科学と技術. 2004, vol. 54, no. 11, p. 568-574.
- 7) 村上晴美, 上田洋. Subject World 新機能: NDC検索, 書籍検索, Web検索. 情報学: *Journal of Informatics*. 2004, vol.1, no.1. <http://kiyo.info.gsucc.osaka-cu.ac.jp/JI/article/view/4/4>, (accessed 2015-05-20).
- 8) Murakami, Harumi; Hirata, Takashi; Kita, Katsuchi. “Subject World: a system for visualizing OPAC”. 人文科学とコンピュータシンポジウム論文集. 大阪, 2002-9-20/22, 情報処理学会人文科学とコンピュータ研究会, 2002, p. 237-240.
- 9) 松本聖, 谷口祥一. “NDC Finder: 自由語からの主題検索機能を提供する図書館 OPAC 検索支援システム”. 2010年日本図書館情報学会春季研究集会発表要綱. 京都, 2010-05-29, 日本図書館情報学会, 2010, p. 79-82.
- 10) 谷口祥一. BSH/NDLSH と NDC における階層構造はどの程度一致するのか. *日本図書館情報学会誌*. 2014, vol. 60, no. 1, p. 1-17.
- 11) Chan, Lois Mai; Zeng, Marcia Lei. Ensuring interoperability among subject vocabularies and knowledge organization schemes: a methodological analysis. *IFLA Journal*. 2002, vol. 28, no. 5/6, p. 323-327.
- 12) Zeng, Marcia Lei; Chan, Lois Mai. Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Sciences and Technology*. 2004, vol. 55, no. 5, p. 377-395.
- 13) Vizine-Goetz, Diane; Hickey, Carol; Houghton, Andrew; Thompson, Roger. Vocabulary mapping for terminology services. *Journal of Digital Information*. 2004, vol. 4, no. 4. <http://journals.tdl.org/jodi/index.php/jodi/article/view/114/113>, (accessed 2015-05-20).
- 14) Vizine-Goetz, Diane; Houghton, Andrew; Childress, Eric. Web services for controlled vocabularies. *Bulletin of the American Society for Information Science and Technology*. 2006, vol. 35, no. 5. http://www.asis.org/Bulletin/Jun-06/vizine-goetz_houghton_childress.html, (accessed 2015-05-20).
- 15) 市瀬龍太郎. 情報の意味的な統合とオントロジー画像. *人工知能学会誌*. 2007, vol. 22, no. 6, p. 818-825.
- 16) 市瀬龍太郎. オントロジー学習: 計算機によるオントロジー構築. *電子情報通信学会誌*. 2009, vol. 92, no. 9, p. 791-795.
- 17) Euzenat, Jérôme; Shvaiko, Pavel. *Ontology Matching*. Springer, 2007, 333 p.
- 18) 国立国会図書館編. J-BISC. 明治期～2005.3. 日本図書館協会, 2006. (DVD).
- 19) 国立国会図書館編. J-BISC. 2005.4～2007.3. 日本図書館協会, 2007. (CD-ROM).
- 20) 国立国会図書館編. J-BISC. 2007.4～2009.3. 日本図書館協会, 2009. (CD-ROM).
- 21) 国立国会図書館編. J-BISC. 2009.4～2011.3. 日本図書館協会, 2011. (CD-ROM).
- 22) 国立国会図書館編. JM-BISC. 2011.4～2013.3. 文字・活字文化推進機構, 2013. (CD-ROM).
- 23) 国立国会図書館編. JM-BISC. 2013.4～2014.3. 文字・活字文化推進機構, 2014. (CD-ROM).
- 24) 地名に細目が付いた件名標目を地名ではなく、普通件名としているものが少数存在した。これらは本来、地名に区分されるべきものであるが、ここではそのまま普通件名として処理した。
- 25) 国立国会図書館収集書誌部. 国立国会図書館件名作業指針. 国立国会図書館収集書誌部, 2013, 209 p. <http://www.ndl.go.jp/jp/library/data/kenmeimanual.pdf>, (accessed 2015-05-20).

要 旨

【目的】 国立国会図書館件名標目表 (NDLSH) において、細目をもたない普通件名など、ある範囲の件名標目に対して、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「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、NDLSH が示す代表分類記号が付与されている数および割合を調査する。

【方法】 1997 年から 2014 年 3 月分までの書誌レコードから、NDC 新訂 9 版分類記号と NDLSH 件名標目の組み合わせをすべて抽出し、それらを普通件名と地名による件名に分けて集計した。また、各件名標目の代表分類記号は、Web NDL Authorities を用いて取得した。

【結果】 1) 普通件名の場合、代表分類記号をもつ件名標目は全体の 20% 以下にとどまり、その殆どは細目をもたない。書誌レコードに付与された分類記号との照合では、件名単位で 80% 以上、件名・分類記号ペア単位で 10% 強が代表分類記号と完全一致となった。分類記号の前方一致まで含めたときには、件名単位でもペア単位でも一致となるものが増加した。対象をレコード内先頭出現の件名標目に限定したとき、一致数および平均一致率の大幅な上昇が見られた。それに対して、細目を除去し主標目のレベルですべての件名標目を照合したところ、これらの値はむしろ低下した。2) 分類記号の側から集計したときには、代表件名標目をもつ分類記号は 30% 強であった。書誌レコードに付与された件名標目との照合では、分類記号単位で 90% 弱、分類記号・件名ペア単位では 10% 強が完全一致となった。継承一致という独自の照合を導入すると、一致となるものが相当数増加した。3) 代表分類記号をもつ地名件名は僅か 6% にとどまり、件名・分類記号のペア単位でも 12% にとどまる。これらのうち、完全一致となるのは、件名単位で 50% 強、ペア単位で 15% 程度であった。