

2 数的組による主題標目の構成

Dyadic Organization of Subject Headings

安 西 郁 夫

*Ikuo Anzai*

*Résumé*

In this article the dyadic organization of subject headings means composing subject headings of descriptors in pair. From the view-point of dyadic organization the writer examines and compares such systems as Medical Subject Headings, Permuterm Subject Index to Science and Technology and SYNTOL, trying to clarify the characteristics of subject headings in each.

The writer then explains about the nature of his own system called CONCODE (Contextual Connection of Descriptors) which combines descriptors by conjunctive marks and prepositions. The dyadic organization principles are applied to CONCODE, providing headings with inclusiveness and CONCODE with exclusiveness.

Although in information retrieval recall and relevance factors are said to act contrary, the writer tries to raise the two factors simultaneously by the combined use of dyadic headings and CONCODE.

(School of Library and Information Science)

- I. 序 説
- II. Medical Subject Headings
- III. Permuterm Subject Index
- IV. SYNTOL
- V. CONCODE
- VI. 結 語

## 2 数的組による主題標目の構成

### I. 序 説

平面上の点は、 $x$  座標と  $y$  座標という 2 数の組によってその位置を示すことができる。点 P の  $x$  座標を  $x_i$ 、 $y$  座標を  $y_i$  とすれば、点 P は、

$$P(x_i, y_i)$$

によって表わされ、同じく点 Q は

$$Q(x_j, y_j)$$

によって表わされる。(図 1 参照)

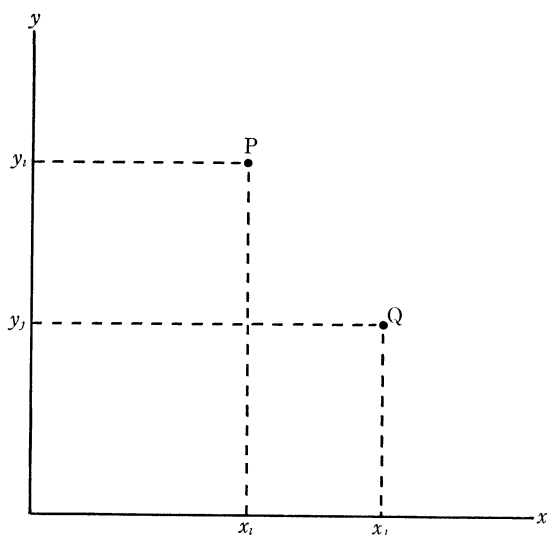


図 1. 座標平面上の点

仮に点 P と Q を主題  $S_i$  および主題  $S_j$  と考え、 $x$  座標と  $y$  座標が記述子 (descriptors) であると見なせば、2 つの主題は次のように表わすことができる。

$$S_i(x_i, y_i)$$

$$S_j(x_j, y_j)$$

このように 2 個の記述子の組による主題表現を筆者は dyadic organization of subject representation と呼んでいる。この方式が、索引の対象となる文献の主題はすべて座標平面上にあり、換言すれば、座標平面は関心の対象たりうる主題の集合であるという前提を持っていることは言うまでもない。

本稿では、検索という道程に立てられる第 1 の道しるべの役割を演ずる subject headings を 2 数的組によって構成する方式について論ずる。Subject headings に対して、図書館界では一般に件名標目という訳語を用いているが、この文脈における subject は必ずしもいわゆ

る“件名”には限られないので、筆者はあえて“主題”という語を用いる。Heading については、主としてドキュメンタリストが“見出し”という訳語を使用しているが、本稿では一応“標目”と呼ぶことにする。

### II. Medical Subject Headings

Sears や BSH などの一般的な件名方式では、主題標目は  $n$  個 ( $n \geq 1$ ) の記述子(自然言語に属する)によって構成される。 $n$  個の記述子の並び方は  $n$  の階乗に等しく、従って、並び方に規制を加えないとすれば、 $n!$  に等しい枚数の件名カードを必要とする。ファイルの作成・保管にかかわるコストの増大を嫌忌する伝統的方式では、一般に標目の数を最小限に留める。

例えば、BSH は次のように規定している。

“件名標目は必要に応じて与えるもので、常に一個とは限らない。二つ以上を与える図書もあれば全然与えなくてよい場合もある。

a. 一主題の図書には一件名標目を原則とする。”<sup>(1)</sup>

このような抑止的規定に加えて、1 標目を構成する記述子の配列にも規制があり、1 番目の記述子(主標目)とそれに続く記述子(細目)の並び方は固定される。例えば、

日本文学——歴史——江戸時代

という並び方が定められると、それは決して

江戸時代——日本文学——歴史

と並び変えられることはない。換言すれば、それはすべての順列の中から 1 つを選択する方式であるということができよう。

前述の如く、記述子は 1 個の場合もあれば、複数個の場合もあるので、必ずしも 2 数的組によって構成されるものではない。例えば 3 個の組によって構成される場合もあるが、その場合においても、主題が 3 次元において捕えられているとは言い難い。仮に 3 個の記述子を  $d_1$ 、 $d_2$ 、 $d_3$  とすれば、 $d_1$  と  $d_2$  は座標平面 I の  $x$  座標と  $y$  座標であり、その組 ( $d_1, d_2$ ) を複合記述子  $d'$  で置きかえるならば、 $d'$  と  $d_3$  とは座標平面 II の  $x$  座標と  $y$  座標となるのである。すなわち、そこには 3 次元の構成があるのではなく、2 次元的操作が 2 回繰り返されているにすぎないと考えられる。

件名方式の中で、2 数的組の構成を原則とするものに *Medical Subject Headings* (MeSH) がある。これは National Library of Medicine (NLM) によって絶え

ず改訂されている生物・医学文献用の索引言語であり、同館編の *Index Medicus* が MeSH を典拠として索引されていることはよく知られている。

MeSH の標目は主標目 (main headings) と副標目 (subheadings) からなっている。1968 年版<sup>2)</sup> では、主標目は A から N までのカテゴリに区分され、そのうち 7 つのカテゴリはさらに 62 のサブカテゴリに細分されている。副標目は新たに追加された 7 個を含めて 60 個用意されている。*Index Medicus* では、1 個の主標目と 1 個の副標目を組み合わせた形で索引することを原則としている。一時副標目の使用が中止されたこともあったが、現在では前述の 2 数的組を用いる方式が復活している。

副標目には、組の相手となる主標目のカテゴリが指定されており、そのカテゴリ数は、最も少ないもので 1 個、最も多いものは 12 個に及んでいる。表 1 は、結びつきうるカテゴリ数別に、該当する副標目を 1 例ずつ示したものである。これを見ると、一般性が強く、専門性

表 1. 結合するカテゴリの個数別に見た標目

順位	結合するカテゴリの個数	副標目 (各 1 例)
1	12	History
2	9	Classification
3	7	Standards
4	5	Instrumentation
5	4	Drug Effects
6	3	Cerebrospinal Fluid
7	2	Anatomy & Histology
8	1	Abnormalities

の稀薄な History, Classification, Standards などが、多くの主標目と組を作ることが明らかである。

次に、主標目の立場から、パートナーとなりうる副標目の数を調べてみると、表 2 の結果が得られる。ここでは、表 1 とは逆に、中心領域のカテゴリは多くの結合相手を持っているが、領域が中心から遠去かるほど結合相手の数は減少している。

主標目に 14 個のカテゴリがあり、各カテゴリには組となる相手の数が指定されているということは、14 個

表 2. 結合する副標目の個数別に見たカテゴリ

順位	カテゴリ記号	カテゴリー	結合する副標目の個数
1	C	Diseases	28
1	F	Psychiatry and Psychology	28
3	A	Anatomical Terms	22
3	D	Chemicals and Drugs	22
5	B	Organisms	14
6	E	Analytical, Diagnostic and Therapeutic Technics and Equipment	10
6	G	Biological Sciences	10
8	H	Physical Sciences	9
9	N	Health Care	7
10	J	Technology, Industry and Agriculture	6
11	I	Anthropology, Education, Sociology and Social Phenomena	5
11	L	Communication, Library Science and Documentation	5
13	K	Humanities	1
13	M	Named Groups of Persons	1

の座標平面 A~N があり、例えば座標平面 A においては、カテゴリ A に属する主標目語が x 軸上に並んでおり、y 軸上には 22 個の副標目語が並んでいると見なすことができよう。

NLM が同館で受入れる図書・雑誌の件名目録を作成する場合には、MeSH に若干の修正を加え、次に掲げる副標目を使用している。

- Form Subheadings
- Geographic Subheadings
- Language Subheadings
- Additional Topical Subheadings

### III. Permuterm Subject Index

Institute for Scientific Information (ISI) によって 1963 年に創刊された *Science Citation Index* (SCI)

## 2 数的組による主題標目の構成

は、引用索引誌としては世界で最初の本格的な商業出版物として斯界の注目を惹いたが、主題アプローチによる検索の用具としての有用性については疑問があった。このような疑問の解消をはかる意味で、1967年秋にISIはSCIの*Source Index*に対する年刊の主題索引として*Permuterm Subject Index to Science and Technology*を発刊した。同索引については、1967年の10月に早くも裏田和夫<sup>3)</sup>が紹介しているが、まだ新しく、しかも高価な索引であるために、わが国ではよく知られていない。

近年もてはやされている索引形式にKWICがある。電子計算機で編集するKWICでは、標題を構成する語のうち、不用語(stop words)を除いた重要語(keywords)を循環させて標目とする。

The kw<sub>1</sub> of kw<sub>2</sub> by kw<sub>3</sub> in kw<sub>4</sub>.

仮に前記のような構造を持つ標題があるとすれば、KWIC索引では、次のような形式で標目が作られる。

- ① KW<sub>1</sub> OF KW<sub>2</sub> BY KW<sub>3</sub> IN KW<sub>4</sub>\* THE
- ② KW<sub>2</sub> BY KW<sub>3</sub> IN KW<sub>4</sub>\* THE KW<sub>1</sub> OF
- ③ KW<sub>3</sub> IN KW<sub>4</sub>\* KW<sub>1</sub> OF KW<sub>2</sub> BY
- ④ KW<sub>4</sub>\* THE KW<sub>1</sub> OF KW<sub>2</sub> BY KW<sub>3</sub> IN

1 標題中に  $n$  個の重要語があれば、車輛の順番を1つずつずらした列車が  $n$  本編成されるのである。

*Permuterm Subject Index* も KWIC と同系の *permutated title index* の 1 種ではあるが、前例の標題は次のような標目形式で処理される。

KW<sub>1</sub>  
 KW<sub>2</sub>  
 KW<sub>3</sub>  
 KW<sub>4</sub>  
 KW<sub>2</sub>  
 KW<sub>1</sub>  
 KW<sub>3</sub>  
 KW<sub>4</sub>  
 KW<sub>3</sub>  
 KW<sub>1</sub>  
 KW<sub>2</sub>  
 KW<sub>4</sub>  
 KW<sub>4</sub>  
 KW<sub>1</sub>  
 KW<sub>2</sub>  
 KW<sub>3</sub>

KWIC の連結が横型であるのに対して、*Permuterm* のそれは縦型であるとも言えるが、実際には2輛編成の列車が  $n(n-1)$  本用意されると見るべきであろう。すなわち、前掲例では、(KW<sub>1</sub>, KW<sub>2</sub>), (KW<sub>1</sub>, KW<sub>3</sub>), (KW<sub>1</sub>, KW<sub>4</sub>), ……(KW<sub>4</sub>, KW<sub>2</sub>), (KW<sub>4</sub>, KW<sub>3</sub>) という2数的組によって構成される標目が  $4 \times 3 = 12$  個作られるのである。括弧中の左側の KW は primary term, 右側の KW は co-term と呼ばれる。

*Index Medicus* における標目の現われ方も、その形式は *Permuterm* と全く同じである。例示すると次のようになる。(MH=main heading, Sh=subheading)

MH<sub>1</sub>  
 Sh<sub>1</sub>  
 Sh<sub>2</sub>  
 Sh<sub>3</sub>  
 MH<sub>2</sub>  
 Sh<sub>1</sub>  
 Sh<sub>4</sub>  
 Sh<sub>5</sub>  
 Sh<sub>6</sub>

以上の例示でも明らかのように、両者はその形式を同じくしてはいるものの、性格は異にしている。すなわち、*Index Medicus* では、MH (主標目) と Sh (副標目) は別物であり、Sh が MH に変換されることはない。ところが *Permuterm* においては、すべての KW が primary term (主標目に相当) にもなれば、また同時に co-term (副標目に相当) にもなるのである。このことを座標平面に即して言うならば、 $x$  軸に並んでいるものと同じ KW が  $y$  軸にも並んでいるということになる。

## IV. SYNTOL

SYNTOL<sup>4)</sup> (*Syntagmatic Organization of Language*) は French National Center for Scientific Research の J. C. Gardin が EURATOM との契約に基いて開発した man-machine system である。

Gardin は、主題表現に関して、“構造パターンはすべて最小の成分、すなわち2数的紐帯に分解できる”<sup>5)</sup> と言う立場をとっている。2数的紐帯(dyadic strings)とは、構文法的に関係する索引語の pair であり、彼はこの pair を syntagma と呼んでいる。

Syntagma の概念を明らかにするために、Gardin が

用いた図形<sup>2)</sup>に基づいて説明を試みよう。

〔例 1〕  $C$  に対する  $B$  の作用を  $A$  が抑止する。

この例では、 $B$  と  $C$  の間に  $R_i$  なる関係があり、その作用-被作用関係に対して  $A$  が  $R_i$  なる関係を持つ。その構造は図 2.1 のようになる。SYNTOL ではこの

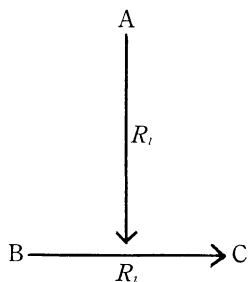


図 2.1

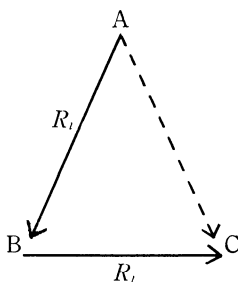


図 2.2

関係を次のように言い改める。すなわち、 $C$  に対して  $R_i$  なる関係を持つ  $B$  に対して  $A$  が  $R_i$  なる関係を持つ——と。その構造は、従って、図 2.2 が示すものとなり、 $R_i(A, B)$ 、 $R_i(B, C)$  という 2 個の syntagma が作られる。

〔例 2〕  $A$  の  $C$  に対する作用と、 $B$  の  $D$  に対する作用の比較

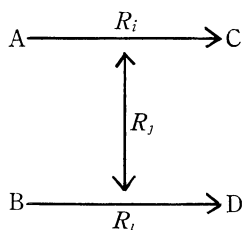


図 3.1

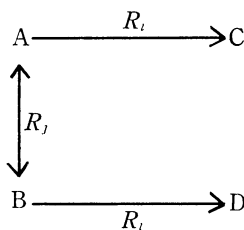


図 3.2

図 3.1 に見られる関係は、 $C$  に対して作用する  $A$  と、 $D$  に対して作用する  $B$  との比較と解釈され、図 3.2 の形に変換される。この場合に作られる syntagma は  $R_i(A, C)$ 、 $R_i(B, D)$ 、 $R_j(A, B)$  の 3 個である。

〔例 3〕 薬品  $A$  は器官  $B$  を刺戟する。

図 4.1 の関係は図 4.2 に変換され、薬品  $A$  を 1、刺戟を 2、器官  $B$  を 3 と置きかえれば、 $R_i(2, 1)$ 、 $R_j(1, 3)$ 、 $R_j(2, 3)$  という 3 個の syntagma が構成される。

以上の 3 例はいずれも単純化されたモデルであり、実際に処理される主題ははるかに複雑ではあるが、複雑な

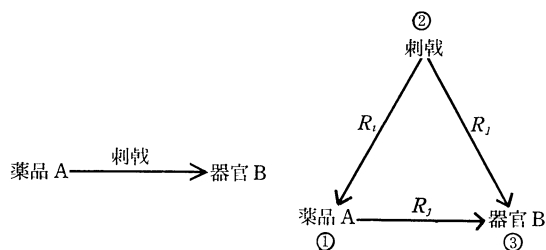


図 4.1

図 4.2

主題も結局は多数個の syntagma に分解される、と Gardin は見なしているのである。

ここで問題となるのは、前掲例において  $R_i$ 、 $R_j$  とし示された構文法的関係であろう。Gardin は次に掲げる 4 種の関係を設定している。

- $R_1$ .....Predicative (叙述的)
- $R_2$ .....Associative (連合的)
- $R_3$ .....Consecutive (連続的)
- $R_4$ .....Co-ordinative (同格的)

これら 4 種の前掲の例 3 に適用すると、次のようになる。

- $R_2(2, 1)$
- $R_3(2, 3)$
- $R_3(1, 3)$

しかしながら、これら 4 種の実際に適用するに当っては、解釈上の混乱が起ることが予想される。特に  $R_2$  と  $R_3$  の適用について問題があると考えられるが、いずれにせよ、2 数的組である syntagma のマトリックス表示を検索に用いるというその構想は興味深い。

## V. CONCODE

筆者は第 4 回ドキュメンテーション研究集会において、「ディスクリプターの文脈的結合による主題目録の一方式<sup>7)</sup>」と題する発表を行ない、CONCODE (Contextual Connection of Descriptors) と呼ぶ方式を提唱した。

これは、括弧、接続記号 &、ハイフン、英語形前置詞により記述子を結合して標題的なフレーズを作り、主題表現体とする方式である。

例えば、(1) 高圧放電による金属板の接着、(2) 春子の和服と秋子の洋服の比較——という主題は次のように表現される。

## 2 数的組による主題標目の構成

- (1) 接着 of 金属板 by 高压 - 放电  
 (2) 比較 of (和服 of 春子) with (洋服 of 秋子)

結合される記述子は規制された語であることを原則とするが、“比較”のような一般用語は、典拠リストに含まれていない場合でも、主題を表現するのに必要であれば、随意に補充語として使用することが許される。前置詞と接続記号は添字のように小さく記入し、記述子を際立たせる。

前置詞は、それ自体の意味内容よりは、第1類品詞を結び合わせる機能を主とする語類である。他品詞から転成した前置詞(第2類前置詞)を除く本来の前置詞(第1類前置詞)は約70語あるが、複合形のものを除く単純前置詞の数は28個である。これらの前置詞は、by, for, in, of, on, to などに見られるように、概してその語形が短かく、外国人の我々にも馴染が深いので、連結子として使用するのに適していると考えられる。

この連結子が、いわゆるリンクの機能を持つことは言うまでもない。例えば前掲第2例の「春子の和服と秋子の洋服の比較」の場合、下線の4語の単純な coordination では、「春子の洋服と秋子の和服」という false combination が作られる。リンクを用いれば、この種の誤りは防げるが、ルールを併用しない限り、関連のある用語間に誤った関連をつけることがある。CONCODE が後者の誤りをも防ぎうるのは、前置詞が被支配語の性格を示す機能を併せ持つからに他ならない。

検索に際して適合文献に到達するためには、我々は通常次の過程を経なければならない。

- 第1ステップ……………Key  
 第2ステップ……………Identificator  
 第3ステップ……………Abstract  
 第4ステップ……………Document

Rolf Moore<sup>9)</sup> は検索の過程を guide と identificator の2つのステップに分けて考えている。筆者の第2ステップの各称は Moore から借用したものであるが、CONCODE はこの第2ステップの identificator の役割を担うものであり、Moore の別の表現に従えば、topic-identificator に相当する。

第2ステップに属する CONCODE は当然その前のステップである key を必要とする。この key の役割を果すものが主題標目である。

第1ステップの key に対して要求されるものは網羅力である。換言すれば、key は主題構成要素のいずれか

らのアプローチをも可能にするものでなければならぬ。第4回ドキュメンテーション研究会における発表では、CONCODE を構成する記述子をそれぞれ単独に標目としてとるとというのが筆者の案であった。仮に  $n$  個の記述子からなる主題があるとすれば、記述子の並び方は  $n!$  通りある。記述子の train をそのまま標目にとり、しかも網羅性をもたせるとすれば、 $n!$  個のエントリーが必要となる。筆者は、標目と CONCODE の2段構成によってそれを  $n$  個に抑え、しかも CONCODE には排他性を付与することにより、情報検索において相反的に作用するといわれる再現率と適合率とを共に高めることを意図したのである。

単一の記述子を標目とする場合に予想されることは、その記述子が使用頻度の高い用語であれば、同一標目のエントリーがきわめて多数となり、選別のために通覧しなければならぬ CONCODE の数が増え、能率が低下するということである。このような事態を防ぐためには、標目のとり方を次のように修正することが考えられる。

### 〔例1〕

CONCODE 接着 of 金属 by 高压 - 放电  
 標 目 (1) 接着 of 金属片  
 (2) 金属片, 接着 of  
 (3) 高压 - 放电  
 (4) 放电, 高压  
 \* 接着 by 放电

### 〔例2〕

CONCODE 比較 of (和服 of 春子) with  
 (洋服 of 秋子)  
 標 目 (1) 和服 of 春子  
 (2) 春子, 和服 of  
 (3) 洋服 of 秋子  
 (4) 秋子, 洋服 of

前掲例において明らかなように、CONCODE を構成する記述子は1回ずつ主標目の座に置かれ、その主標目と最も連合度の高い記述子が副標目として並べられる。CONCODE の中で隣りあう記述子同志の連合度が高いのが通常であるが、問題は隣り合わない記述子との間にも強い連合度が存在する場合であろう。このような場合、標目数を増やしうる状況にあるならば、例1の\*印のような標目を追加すればよい。

いずれにせよ、この方式では、標目は2数的組によって構成される。同じ記述子の組を標目とするエントリー

の配列順をどうするかの問題が残されているが、これは記述子間に介在する連結子に順位を与えることによって解決できる問題である。ハイフンやコンマを優先させ、接続記号と前置詞にはアルファベットの順位を与えればよいであろう。

## VI. 結 語

2 数的組による主題標目の構成という観点から MeSH, Permuterm, SYNTOL ならびに CONCODE を順次論じて来たが、情報検索の key となる標目については、新しい観点から多角的な考察を加える必要があるであろう。2 数的組による構成に関しては、その実際の効率を多面的に測定する比較の実験が行なわれない限り、それが他に比して優れていると即断することはできない。しかしながら、この種の実験には多くの困難が伴うので、非力な筆者には荷が重すぎるようである。せめて筆者の方式に対して諸先輩から忌憚のない御批判を頂ければ幸いである。

(図書館・情報学科)

- 1) 日本図書館協会件名標目委員会. 基本件名標目表. 同会, 1956. p. 32.
- 2) U. S. National Library of Medicine. *Medical subject headings*. Washington, D. C., U. S. Public Health Service, 1968. 413 p.
- 3) 裏田和夫. "Science Citation Index—Permuterm Subject Index to Science and Technology. 1966—Philadelphia, Institute for Scientific Information," *Library system*, vol. 6, 1967. 10, p. 275-8.
- 4) Gardin, J. C. *SYNTOL*. New Brunswick, N. J., Graduate School of Library Service, Rutgers the State University, 1965. 106 p. (Rutgers series on systems for the intellectual Organization of information, vol. 2)
- 5) *Ibid.*, p. 27.
- 6) *Ibid.*, p. 28-31.
- 7) 安西郁夫. ディスクリプターの文脈的結合による主題標目の一方式. <第4回ドキュメンテーション研究集会, 1967. 発表論文集. 日本科学技術情報センター, 1968> p. 69-73.
- 8) Moore, Rolf. 前園主計訳. "ドキュメンテーションにおける identifier の原理," *ドキュメンテーション研究*, vol. 18, no. 5, 1968. 5, p. 138-40, 47.