

シソーラスの問題点

On Some Problems of Thesauri

安倍浩二

*Koji Abe*

*Résumé*

The problems of thesauri for document retrieval have been discussed in many papers so far, and guidelines for the establishment of thesauri have already been published by UNESCO and ISO.

The guidelines are very insufficient concerning the purpose and criteria for preferential cross-referencing between the descriptor and the non-descriptor, and for hierarchical and associative cross-referencing between descriptors.

The author thinks that this insufficiency is due to lack of consideration of what documents are necessary to get the answer to a request, including the term or the concept discussed. In these guidelines only synonymous, generic and part-whole relations between terms are mentioned as basis for preferential and hierarchical cross-references.

In spite of the lack of clear description of this viewpoint in the guidelines, in various existing thesauri cross references between terms are made, on the basis of relations between documents.

For example, the Engineers Joint Council thesaurus, which restricts hierarchical cross-references primarily to generic relations, adopts relations other than synonymous and generic ones as basis for preferential cross-referencing.

Reviewing the above-mentioned guidelines and examples in existing thesauri in terms of definition of thesaurus, selection of descriptors, cross-references between terms, special procedures for generic terms, and methods of compiling thesaurus, the author discusses specific problems and suggests some solutions.

- I. シソーラスの定義
- II. ディスクリプタの選定基準
- III. ディスクリプタ, 非ディスクリプタの間の関係

---

安倍浩二：日本科学技術情報センター情報部主任情報員  
Koji Abe, Senior Information Officer, Information Division, The Japan Information Center of Science and Technology.

- A. 優先関係と階層関係
  - B. 関連関係
- IV. 一般語の取扱い
- V. シソーラスの作成
- A. フリータームによる索引作業の結果から人手により作成する方法
  - B. フリータームによる索引作業の結果から機械的に作成する方法
  - C. 専門研究者に索引作業と同時に語の関係を作成させ、電子計算機などによりシソーラスを作成する方法
  - D. 外国語シソーラスの翻訳
  - E. 既存の分類表、件名標目表、シソーラスにもとづく方法
  - F. 既存の用語集、専門家の知識をもとにして選択された基本的な語の階層関係、関連関係をきめながら、シソーラスを作成する方法

## I. シソーラスの定義

UNESCO の guideline<sup>1)</sup> には次のように書いてある。

“シソーラスは機能あるいは構造の上から定義することができる。

機能の上からは、シソーラスは文献、索引作業者あるいは利用者の自然言語をより統制された「システム言語」（ドキュメンテーション言語、情報言語）に翻訳し、あるいは逆に、システム言語を自然言語に翻訳するために使用される用語統制手段である。

構造の上からは、シソーラスとは知識の専門領域を包括的にカバーする意味的およびクラス・メンバー的に (generically) 関連のある term の統制された動的な (dynamic) 語い (vocabulary) である。この語いはディスクリプタ、非ディスクリプタとそれらの関係の組織的およびまたは alphabetical な collection である。”

Wersig の定義<sup>2)</sup> によれば、

“ドキュメンテーション領域におけるシソーラスは、概念の分類および秩序的ための記号の秩序化された集合である。分類システムとしてシソーラスは記号の概念への可逆的—義的秩序付けを旨とし、秩序化のシステム (Ordnungssystem) としては概念間の関係の記号化を旨とする。このシステムは概念の分類と秩序化のための開かれた (offene) システムである。”

VINITI の「情報科学技術辞典」<sup>3)</sup> では以下のように定義している。

“情報検索言語の語いを構成する単位（ディスクリプタ）の全部をそれらの間の関係、同義ディスクリプタ、キーワードおよび自然語の句と共に列挙した辞書型便覧”

この定義中のディスクリプタの意味は索引言語（記号を含む）の語い構成単位一般の意味であろう。

これらの3つの定義は手段あるいは語い、記号の集合、辞書型便覧という風に表現は違っているが、これらの3つの定義を総合すればシソーラスの機能と構造が明らかになる。term あるいはディスクリプタ間の関係と考えるか、2番目の定義のように概念間の関係の記号化されたものとするかは、違うようであるが、term あるいはディスクリプタを概念を表現するものと考えれば同じことになる。

UNESCO の定義の dynamic と2番目の定義の offene は同じだと思われる。後者では offene System を外部の環境との交流が行なわれるシステムと定義していることから、技術の進歩に伴ってたえず更新されること、すなわち外部の環境からのフィードバックがたえず行なわれることを意味していると思われる。

シソーラスは索引作業と検索の実用的なツールである。研究用に試作されたシソーラスはここでは考えない。実用的なツールであるためには「使いやすい」ということが必要条件である。そのためには理論的な一貫性のある程度ぎせいにしなければならない。研究用に試作されたシソーラスは比較実験などのため理論的な一貫性

を要求されることが多い。

しかし実用シソーラスといえども、理論あるいは作成規則からのずれがあまりに甚しいと逆に使いにくくなる。

## II. ディスクリプタの選定基準

シソーラスを作る上でもっとも重要なことはディスクリプタの選定基準である。それで別に新しい問題ではないが問題点として検討して見る。

まず参考として UNESCO の guideline<sup>4)</sup> の 5.5 selection of descriptors の項の翻訳を示す。

### “(5.5) selection of descriptors

一般に、ディスクリプタの選定には、collection, verification, evaluation, choice のはっきり分れた4つの段階がある。”

#### (5.5.1) collection

ABC 順に候補語を網羅的に集めることはほとんど不可能である。ディスクリプタをグループに分けて考察することにより、それらのディスクリプタ間の関連から多くの候補を選ぶことができる。語を集めるためのソースには次のようなものがある。

利用者となるべき人および各分野の専門家

国際的または国別の標準辞書

現在配布されている文献

術語に関する論文または術語表

現存のシソーラスおよび分類表

命名法

雑誌の索引

抄録サービス

教科書、手引き書 (handbooks) および摘要書

文献の実験的索引作業

いくつかの分野には命名法の高度に特殊なシステムあるいは十分に確立した技術語が存在する。国際的に承認された命名法がある場合にはそれを使用すべきである。関係付けされていない特殊な名前がふえると、シソーラスを identifier の単純なリストに変え、シソーラスの機能を果さなくなる。したがって、関係付けされていない特殊な実体 (entities) の名称はできるだけ避けるべきである。ディスクリプタとして取扱われない (すなわち、相互の関係を示されていない) 特殊な名称はシソーラスの主部分とは別にリストにしてもよい。

#### (5.5.2) verification

あらゆる方法で討議の会合を開くと共に、選択されたディスクリプタの出所の正しさ (authenticity) を辞書類、その他の索引作業用あるいは標準語い、現在の文献中の使用状況および特に分野の専門家の意見を参照することによって確かめる。

シソーラスの興味ある特性の一つは基礎および応用分野で増大する新造語句使用や特殊な俗称を直ちに吸収する能力である。スコープノートおよび相互参照の使用を併用することによりこの特性の利点を十分に活用すべきである。その言外の意味が時と共に変化した場合、その意味が国により異なる場合、それらの用語には特殊な注意が必要である。重複する (overlapping) 用語を含まなければならない時には適当な相互参照が必要である。

#### (5.5.3) evaluation

ディスクリプタ候補の有用性を評価するには次の事項を参照する必要がある。

- (a) 文献中あるいは現存の情報蓄積中の使用ひん度
- (b) 検索質問中での予期される使用
- (c) 既存のディスクリプタとの関係
- (d) 関連分野の現行語いとしての適切さと出所の正しさ (authenticity)
- (e) 特定の概念の内包と外延を現わす (connote and denote) 上での効果と利便さ

これらの因子は独立に考慮すべきでなく、周辺分野に対しては特別な注意が必要である。それらの領域では網羅性 (exhaustivity) と特殊性 (specificity) に対する要求は中核的な事物 (core subject) における要求と同一ではない。

#### (5.5.4) choice

すべての場合、ディスクリプタは検索の目的に対して予測される効果にもとづいてシソーラスに入れるための選択を行なうべきである。”

上の guideline を参考として、もっと具体的なディスクリプタ選定基準を考えて見る。すなわち

a) 対象とする分野の対象とするすべての文献の索引作業に必要な概念を表現するディスクリプタを網羅していること。

b) 同一の概念を (単独のディスクリプタあるいはその組合せで) 表現する方法は一通りしかないこと。概念は異なっても文献群が同一と考えられる場合は同一の方法で表現しなければならない。(3.1 例3参照)。

c) 1 個のディスクリプタが表現する概念の範囲が明

確であること

d) ディスクリプタの表現は標準的、普遍的であること

e) 検索に使用されることが期待されること

f) ディスクリプタの文献中の出現ひん度、あるいは、そのディスクリプタと同種同レベルのディスクリプタ群の文献中の平均出現ひん度が高すぎもせず低すぎもせずある範囲内にあること

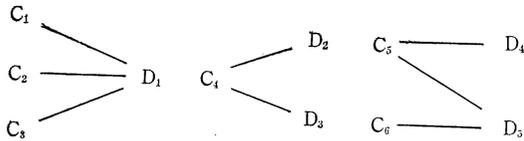
g) ディスクリプタの総数がある限度以下であること

a) の条件は、文献の主題分析により文献に関連ありと判断されたすべての概念を、その specificity のレベルでそのまま表現するディスクリプタがすべて選定されている、ということではない。そのようなことが不可能なことは g) の条件から当然わかる。

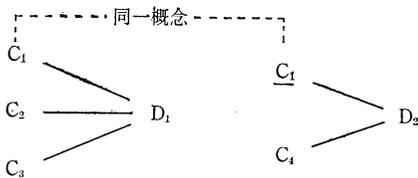
a) と b) の条件から対象文献群から主題分析によって抽出されたすべての概念に対して、各概念に対応する単独のディスクリプタあるいは複数のディスクリプタの組合せが存在するということになる。

すなわち、概念を  $C_1, C_2, \dots$  ディスクリプタを  $D_1, D_2, \dots$  とすると第1図のようになり、第2図のようにはならない。

第1図からわかるように一般に概念とディスクリプタの対応は多対多である。



第1図 概念とディスクリプタの正しい対応



第2図 概念とディスクリプタの正しくない対応

G. Wersig のシソーラスの定義中の「記号の概念への可逆的一義的秩序付け」(die umkehrbar eindeutige Zuordnung von Bezeichnung zu Begriffen) は上のような意味であろう。

c) の条件はディスクリプタそのものの表現、そのディスクリプタと非ディスクリプタあるいは他のディスク

リプタとの関係の表示、スコープノートによって満たされる。以下その各々についてやや詳しく述べる。

(1) ディスクリプタそのものの表現

ディスクリプタの表現は自然語による表現であるから、そのままでは自然語がもつ同義性 (synonymy または synonymity), 同形異義性を含んでいる。同義性は同義語中の1個のみをディスクリプタに選定することによって排除される。同形異義語 (homograph) は、語頭に同形異義を区別するための語あるいは接頭辞を付けるか、それができない場合には [ ] に入れた限定句を語尾に付加するなどの方法によって区別する。

例 変態 → 変態〔生物〕

    \ 相変態または相転移

日本語の場合には漢字を処理する情報蓄積検索システムが普及していないため、かな文字のディスクリプタを用いることが多い。その場合は漢字の場合と比べると同形異義語がいちじるしくふえる。これを防ぐにはなるべく発音の違う語をディスクリプタにすることが望ましい。区別するには日本でも慣用されている外国語で表わすとか、発音の違う日本語にいかえるなどの方法がある。

例 カネツ 加熱 → ヒーティング  
           過熱 → スーパーヒーティング  
       ハイスイ 背水 → バックウォーター  
           廃水 → ハイエキ (廃液)  
       チカク 地核 → アースコア  
           地殻 → アースクラスト

(2) 非ディスクリプタとの関係表示

略語はなるべく使用しない方がよいが、使用上の便利さのため使用したいことがある。その場合にはフルネームまたはその日本語訳を非ディスクリプタとして参照を付ける。

例 1. 「JICST シソーラス 1975 年版」<sup>5)</sup>  
       ディスクリプタ           COD  
       非ディスクリプタ   化学的酸素要求量  
       化学的酸素要求量       USE COD  
       COD UF 化学的酸素要求量

例 2. 「EJC シソーラス (TEST)」<sup>6)</sup>  
       ディスクリプタ   Manuals (手びき)  
       非ディスクリプタ  
           Instruction manuals (指導要領)  
           Maintenance manuals (整備要領)  
           Technical manuals (技術便覧)



献中の出現ひん度であるが、その他の重要な基準は複数ディスクリプタの組合せによる概念の表現をどの程度採用するかである。

JICST シソーラスでは化合物、合金の組合せによる表示については一応の基準ができてはいるが、他の一般の複合概念の表示などについてはまだ基準がない。

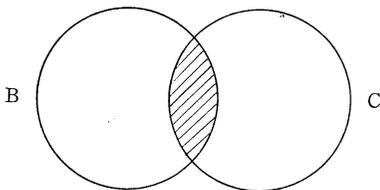
この場合もシソーラスの使いやすさの点から、ある種の複合概念を示すディスクリプタは一括して採否をきめるべきであろう。

たとえば、「鋼板」、「プラスチックフィルム」、「ガラス管」など材料と形の複合語は現在は個々の語の出現ひん度の高いものをディスクリプタとして採用しているが、組合せの数が非常に多く全部ディスクリプタとして採用することは到底不可能なので、全部一括してディスクリプタから除外し、「鋼板」はディスクリプタ「鋼」と「板」を組合せて表現するというように、材料のディスクリプタと形のディスクリプタの組合せで表現するように改めるべきであろう。

以下、概念の複数ディスクリプタによる表現の検索方法、検索効率に関する問題についてやや詳しく述べる。

一つ概念を単独のディスクリプタAで表現した場合と2個のディスクリプタB、Cで表現した場合を考える。

単独のディスクリプタAで表現した場合、Aに関する文献の検索はディスクリプタAによって行われる。2個の場合ディスクリプタB、Cの論理積で検索される。論理積は第3図のベン図の斜線の部分に相当する。ただ



第3図 ディスクリプタB、Cの組合せで表現される文献群

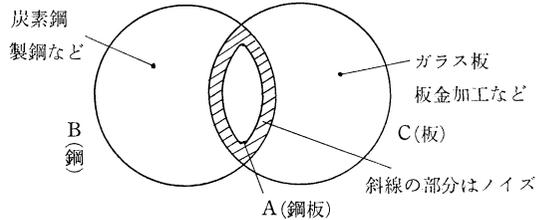
し、このベン図は文献群に関するベン図であって概念に関するベン図ではない。

もしAに関する文献群、すなわち、ディスクリプタAを付与されるべき文献群がこの斜線の部分と完全に一致するならば、ディスクリプタB、Cの論理積によってAに関する文献群がもれもノイズもなく検索されることになる。

しかしコーディネート索引法の場合は必ず若干のノ

イズが生ずる。

検索もれについてはどうであろうか？もしディスクリプタAに関する文献がディスクリプタBに関する文献群の一部であり、かつ、ディスクリプタCに関する文献群の一部でもあれば、B、Cの論理積による検索はノイズを伴っても検索もれを伴わない。すなわち第4図のようになる。



第4図 単独のディスクリプタAと2個のディスクリプタB、Cが表現する文献群の関係

Aが鋼板、Bが鋼、Cが板の場合、索引作業が正しく行なわれていれば第4図のようになるであろう。

すなわち、「鋼板」に関する文献群は「鋼」に関する文献群の一部であると同時に、「板」に関する文献群の一部でもあるように思われる。

Aが「コバルト同位体」、Bが「コバルト」、Cが「同位体」の場合はどうであろうか？「コバルト同位体」に関する文献群が「同位体」に関する文献群の一部であることには異論がなさそうであるが、「コバルト」に関する文献群の一部かどうかには疑問がある。たとえば、コバルト-60はコバルト同位体の一種であり、「コバルト-60によるがんの治療」に関する文献は「コバルト同位体」に関する文献であると同時に「同位体」に関する文献ではあろうが、「コバルト」に関する文献かどうか疑わしい。

この場合「コバルト同位体」に関する文献にはディスクリプタ「コバルト」と「同位体」を付与することにきめ、その通りに索引作業が行なわれていれば、それらの文献群は「コバルト」と「同位体」の論理積でもれなく検索される。しかしディスクリプタ「コバルト」を付与された文献の中には上の理由で「コバルト」に関する文献として疑わしいものが含まれている。したがって「コバルト」に関する文献をディスクリプタ「コバルト」で検索する際にノイズがかなりふえるおそれがある。

しかし検索もれさえなければノイズは、ある程度多くてもいい場合には、こんなことを気にしないで、なるべ

く少数のディスクリプタを組合わせて付与する方がよいことになる。

次の場合にはノイズでなく検索もれが生ずる。すなわち、B、Cの何れかに、たとえばBに対し下位概念に相当するディスクリプタがあるのに階層関係が付けられていない場合である。この場合にはディスクリプタBを付与される文献にはその下位概念に相当するディスクリプタに関する文献が含まれていないから、BとCの論理積による検索ではもれが生ずる。階層関係が付いていればup-posting（上位語付加）あるいは下位語による検索の追加により検索もれを防止できる。

例 ディスクリプタ「化学」の下位概念に相当する「有機化学」、「無機化学」などのディスクリプタが「化学」のNTになっていない場合には、「化学教育」に関する文献を「化学」と「教育」の論理積で検索すれば有機化学、無機化学の教育に関する文献がもれる。

階層関係を持たない孤立ディスクリプタを組合せてindexing する場合にはこのような危険が伴なう。

ディスクリプタBのある観点からの下位概念を表わすディスクリプタを、Bと他のディスクリプタの組合せで表現し、その他のすべての観点からの下位概念はBのNTにすれば、上に述べたような検索もれの心配はない。

たとえば「温度」というディスクリプタの下位概念の内「水温」、「油温」、「排気温度」のような「温度が問題となる物質の種類」と温度の複合概念はすべて物質の種類を示すディスクリプタ（「水」、「油」、「排気」など）とディスクリプタ「温度」を組合せて表示し、「極低温」、「高温」など温度範囲を示すディスクリプタ、「臨界温度」など特定の状態を示す温度など他の観点の下位概念をすべてディスクリプタにしてあれば、「水温」などを「水」と「温度」の組合せで検索してももれは生じない。ただし物質の方も下位概念がNTになっていないものは不適合である。

要するに検索もれを少なくするためには、言い換えれば再現率をよくするためには、ある種の複合概念は一括してディスクリプタから除外してディスクリプタの組合せで表現し、組合せに使用するディスクリプタには下位概念に相当するディスクリプタがある限り階層関係を付けるべきである。

### III. ディスクリプタ、非ディスクリプタ間の関係

実際に問題になるのはディスクリプタと非ディスクリ

プタの間の優先関係とディスクリプタ相互間の階層関係、関連関係などである。

ディスクリプタの代りに記号（概念番号など）を用い非ディスクリプタを自然語にもとづく語とする場合もあるが、その場合にも以下述べることはそのまま成立つ。

以下ディスクリプタと非ディスクリプタを総称して語と呼ぶ。

語間の関係には語の意味あるいは対応する概念の内包あるいは外延によってきまる関係と、語に対応する文献群間の関係がある。

前者は安定した普遍的な関係であり、後者は対象文献の分野、索引作業規則に依存する比較的不安定な関係である。

しかし後者でも対象分野を限定し、索引作業規則をきめることによって安定化することができる。後者をソコロフは語用論(Прагматика)の立場から説明している。<sup>7)</sup> シソーラスの目的は検索の能率化であり、そのためには後者の関係が直接に役に立つ。

#### A. 優先関係と階層関係

##### 1. 優先関係と階層関係を付ける目的

前者はディスクリプタと非ディスクリプタの関係であり、後者はディスクリプタ相互間の関係であるが、両者には共通な点がある。

Nを非ディスクリプタ、D<sub>1</sub>、D<sub>2</sub>をディスクリプタとし、その間に次の関係があるとすると、

優先関係 N USE D<sub>1</sub>

階層関係 D<sub>1</sub> BT D<sub>2</sub>

すなわち、非ディスクリプタNをディスクリプタD<sub>1</sub>によって表現し、ディスクリプタD<sub>2</sub>はディスクリプタD<sub>1</sub>の上位語すなわちbroader term であるとすると、

優先関係にはNとD<sub>1</sub>が同義語、すなわち同一の概念を表示する場合と、NとD<sub>1</sub>が準同義語、すなわち異なる概念を表示するが、索引作業および検索の際NをD<sub>1</sub>で代表させる場合とがある。前者の関係を同義関係、後者の関係を準同義関係と呼ぶことにする。一個のNを複数のディスクリプタの組合せで表現する場合には、Nと各ディスクリプタの間に複数の準同義関係が成立つと考えることができる。

既存のシソーラスの優先関係、階層関係の定義を無視して、純粋に検索の立場から考えて見ると、Nに関する文献群がD<sub>1</sub>に関する文献群に完全に含まれ、D<sub>1</sub>に関する文献群がD<sub>2</sub>に関する文献群に完全に含まれていれば、D<sub>1</sub>で検索すればNに関する文献がもれなく検索さ

## シソーラスの問題点

れ、 $D_2$  で検索すれば  $D_1$  に関する文献がもれなく検索される。

以上、非ディスクリプタ  $N$  とディスクリプタ  $D_1$ 、 $D_2$  について考えたが、一般化して語  $A$  と語  $B$  について考える。

「語  $A$  に関する文献群が語  $B$  に関する文献群に完全に含まれる」というのはどういう意味であろうか？

アンダーラインした「に関する」と「に完全に含まれる」の意味を考えてみよう。

まず、「に関する」とは何か？

ある語に関する文献とは、その語が定義されていれば、ある概念に関する文献である。ある概念に関する文献とはその概念に関する検索質問の答として適当な文献ということで、個々の利用者の関心にしがたがって彼の所属する分野、彼の職種（研究者であるか現場技術者であるかなど）に依存する。しかし特定の検索システムではカバーする分野、利用者の職種の範囲がきまっているものと考えれば、個々の利用者ではなく、検索システムが標準的な対象として認めた分野と、利用者の職種とに依存することになる。

特定のディスクリプタ、非ディスクリプタに関する文献とは、上の語あるいは概念の内シソーラスに登録された語に関する文献で、シソーラスのスコープノート、語間の参照によって限定されたディスクリプタ、非ディスクリプタのスコープに含まれる概念に関する文献、さらにディスクリプタの場合はそれを使用するための規則である索引作業規則に従って実際にそのディスクリプタを付与すべき文献である。シソーラス中の語間の参照の内の優先関係は、この索引作業規則の細則と考えることができる。

たとえば第II章のディスクリプタ選定条件の条件 c) に関連して述べた例

State of the art study USE Reviews は ‘State of the art study’ (研究の現状調査) に関する文献にはディスクリプタ ‘Reviews’ (総説) を付与せよという具体的な指示を与える索引作業の細則である。

実際の索引作業ではある文献に対し該当するもっとも specific なディスクリプタのみを付与する機会が多いが、ここでは階層関係の上位にあるすべてのディスクリプタを付与すると仮定する。

実際の情報検索システムではもっとも specific なディスクリプタのみを付与し、すべての上位ディスクリプタをコンピュータにより索引ファイルに自動付加するか、

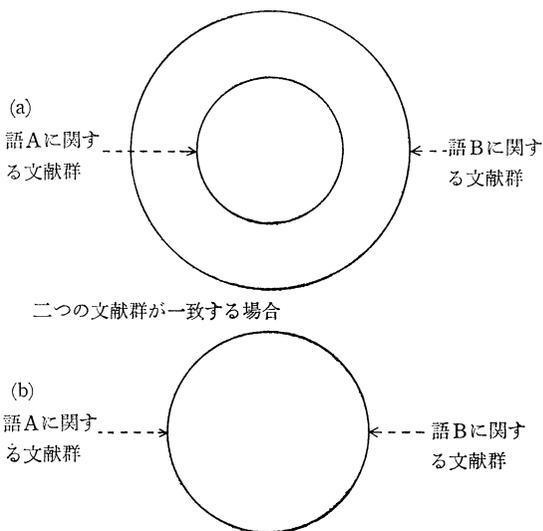
検索の際階層関係のすべての下位ディスクリプタを論理和の形で付加するか、どちらかの方法で上と同じ結果を得ることができる。

以下、条件を単純化するため実際の索引作業がすべての上位のディスクリプタを付与することを含めて完全に索引作業規則に従って行なわれたと仮定する。

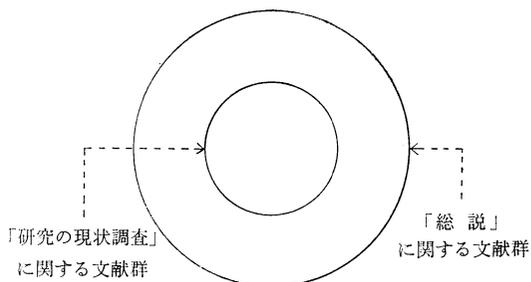
次に、「に完全に含まれる」とは何か？

「語  $A$  に関する文献群が語  $B$  に関する文献群に完全に含まれる」とは、上で定義した「語  $A$  に関する文献群」が「語  $B$  に関する文献群」の部分集合であるということである。部分集合の特殊な場合として二つの文献群が完全に一致する場合も含むこととする。

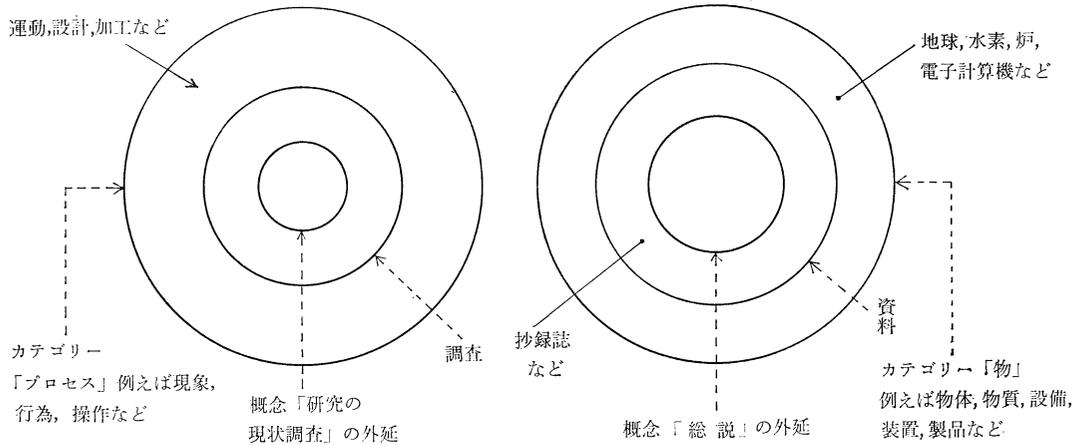
これをベン図で表わせれば第5図のようになる。このベン図も第3図、第4図と同様、文献群に関するベン図である。このような関係を語  $A$  と語  $B$  間の文献包含関係と



第5図 文献包含関係



第6図 文献群間の関係のベン図



第7図 概念の外延間の関係のベン図

呼ぶことにする。

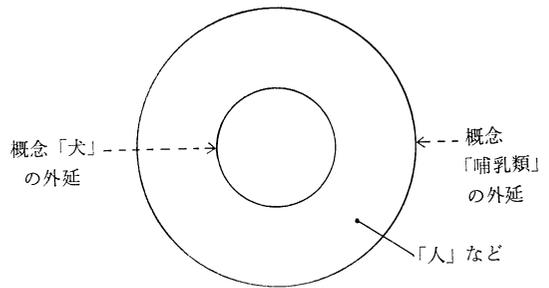
ここで注意しなければならないのは、文献群間の関係である文献包含関係と概念間の関係とは互いに関連はあるが別の関係であるということである。

例 1. 先の例の「研究の現状調査」と「総説」とは文献包含関係にあるとしても、概念としては別のカテゴリーに属する概念でその外延は全く重複しない。これをベン図で表わせれば第6図、第7図のようになる。

例 2. クラスメンバー関係（包摂関係、類概念と種概念の関係あるいは英語の generic relation) は同時に文献包含関係であることが多い。たとえば、概念「哺乳類」と「犬」はクラスメンバー関係にあり、「犬」に関する文献群は「哺乳類」に関する文献群の一部と考えられる。

これをベン図で表わせれば第8図、第9図のようになる。

例 3. 文献包含関係の特殊な場合として二つの文献群が一致する場合。これには二つの場合がある。



第9図 概念の外延間の関係のベン図

a) 概念と文献群が共に一致する場合。これは同義語間の関係である。たとえば「包摂関係」と「クラスメンバー関係」。

b) 概念は一致しないが文献群が一致する場合。たとえば「同義語」と「同義関係」。(同義語間の関係)は概念は一致しないが、「同義語」に関する文献群と「同義関係」に関する文献群は完全に一致すると考えてよいのではないか？

これらをベン図で表わせれば第10図、第11図のようになる。

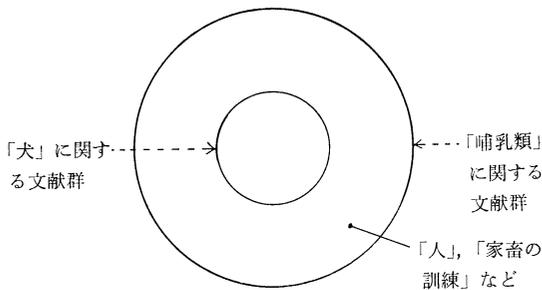
語A, B間の文献包含関係を

$$A \subseteq B$$

で表わす。記号  $\subseteq$  の  $\subset$  は語Aに関する文献群が語Bに関する文献群の一部である場合、 $=$  は二つの文献群が一致する場合を示す。

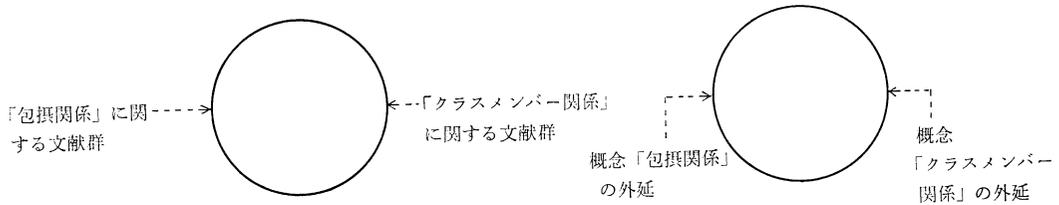
語A, Bの他に語Cを考え

$$A \subseteq B \text{ かつ } B \subseteq C \text{ であれば}$$

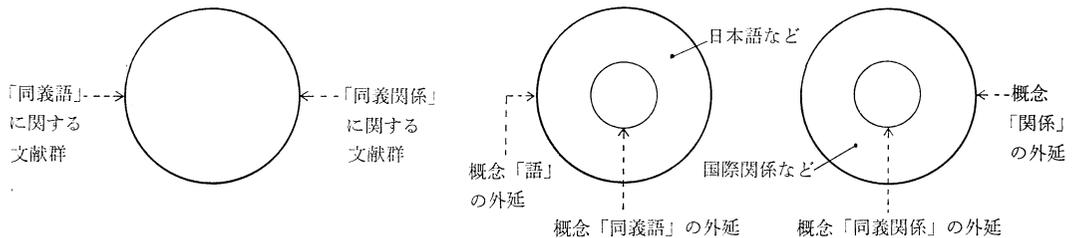


第8図 文献群間の関係のベン図

シソーラスの問題点



第10図 文献群と概念の外延間の関係のベン図



第11図 文献群と概念の外延間の関係のベン図

$$A \subseteq C$$

が成立つ。したがって文献包含関係は推移的關係である。

次に優先関係  $N \text{ USE } D_1$  と階層関係  $D_1 \text{ BT } D_2$  について再び考える。このような関係すなわち参照が付けられているということは、「 $N$ に関する文献群が  $D_1$ に関する文献群に完全に含まれ、 $D_1$ に関する文献群が  $D_2$ に関する文献群に完全に含まれる」ということであると考える。これは、 $N$ と  $D_1$ 、 $D_1$ と  $D_2$ の間に上で定義した文献包含関係が成り立つことである。そして  $N$ と  $D_1$ に関する文献群は一致することがある（同義関係と準同義関係の内の文献群が一致する場合）が、 $D_1$ と  $D_2$ に関する文献群は一致しない（一致すればどちらかをディスクリプタ他を非ディスクリプタにしなければならない）から、

$$N \subseteq D_1$$

$$D_1 \subset D_2$$

となる。そして

$$N \subset D_2$$

も成り立つ。

ただし実際のシソーラスの優先関係、階層関係の中には文献包含関係でないものがあるが、文献包含関係を目標とすべきであると考え。文献包含関係でない場合には検索の際にノイズが生ずる。

2. 既存のシソーラスの優先関係、階層関係の基準

① EJC (米国工学者協会) シソーラス

優先関係には上述の

State of the art study USE Reviews

のような関係がある。

このような関係は行為と物の関係でクラスメンバー関係ではない。

これに対して階層関係は地理的名称や動物の体の部分における全体一部分関係以外は原則としてクラスメンバー関係である。

以下に体の部分とクラスメンバーの関係の例を示す。

例1 Cardiac ventricles (心室) BT Heart (心臓)

例2 Reviews (総説) BT Documents (文書)

② EURATOM (欧州原子力機構) シソーラス、およびこれにもとづいて作成された INIS (国際原子核情報システム) の初期のシソーラス

優先関係も階層関係もクラスメンバー関係以外の関係をかなりの比率で含んでいる。<sup>8)</sup>

優先関係の例

AERIAL SURVEYING (空中探査)

USE AIRCRAFT (航空機)

AND PROSPECTING (探査)

PROSPECTING との関係はクラスメンバー関係であるが、AIRCRAFT との関係はそうではない。

階層関係の例



シソーラスの問題点

通のシソーラスとは異なる。またこのシステムは実験システムなので種々のシソーラスを使ってその検索結果を比較している。

優先関係に当るもののうち、主なものは次の3種類である。

a) word と語幹

例	weld	}	weld
	welded		
	welding		

b) 語と概念コード<sup>11)</sup>

例	概念コード	語
344	obstacle	target
345	atmosphere	meteorology
	weather	wind
346	aircraft	airplane
	bomber	craft
	helicopter	missile
	plane	

a) と b) は実験に使われた二つの独立の方法である。a) は必ず多対一の対応であるが、b) は同形異義の場合、同一の語が複数の概念コードを持ち得る(たとえば、wind は気象関係の概念コード 345 の他に「巻く」という意味の概念コード 233 を持つ)ので対応は一般には多対多である。

c) 概念コードの組合せと概念コード<sup>12)</sup>

例 error を示す 0069 と correct を示す 0306 の組合せ

組合せの概念コード 0280

この場合は組合せの要素となる概念コードは b) によって一定の語と対応するので、b) と c) の併用となる。何れにしても語を非ディスクリプタ、概念コードをディスクリプタと考えた場合、その関係は同義関係、クラスメンバー関係以外の関係を含む。

階層関係に当るものの例として下記がある。

上位概念コード	下位概念コード
0069 (error を示す)	0307 (noise を示す)
0069 (error を示す)	0306 (correct を示す)

0280 (c. の例をみよ)

0207 (verify を示す)

このように階層関係はクラスメンバー関係以外の関係を含む。上の例の中でクラスメンバー関係と見なされ得るのは error と noise の関係だけである。

3. 既存の分類法における相関索引の見出し語と分類記号が示す概念の関係と、分類階層中の分類記号が示す概念との関係

UDC<sup>13), 14)</sup> の例について考える。

優先関係に相当するのは相関索引(逆引表)の見出し語と標数との関係である。

例 1. 情報検索 002.5

分類表の対応する項目を見ると

002 ドキュメンテーション

.5 文書の複製、複写および選出; 機械的手段となっている。情報検索は選出の一種と考えられるから、この場合優先関係に相当するものの意味的關係はクラスメンバー関係である。

例 2. 索引作成 001.815

分類表の対応する項目は

001 学問と知識に関する一般問題

.8 一般方法論

.81 知的作業の技術; 資料収集, 発表の方法

.815 目録, 索引の作成

例 1 と同様優先関係に相当するものはクラスメンバー関係である。

例 3. 一般図書館 027

027 一般図書館

索引と分類表の表現は完全に一致している。同義関係に相当する。

例 4. 意味論(文法) 801.54

801.54 語源学, 語の形成, 意味の変化

「意味論(文法)」と「意味の変化」との関係はクラスメンバー以外の関係である。したがって「801.54」と「意味論(文法)」との関係もクラスメンバー以外の関係である。

階層関係に相当するのは UDC の場合主標数間, 補助標数間の階層関係である。

上の例 1, 例 2 の場合大体においてクラスメンバー関係であるが、下の例で示されるようにそうでないものも多い。

例 5.

003 文字と書法; 記号, 符号, 記号学, 文書

構成法および語義研究一般

## .5 文字を書くための材料と器具

## 例 6.

## 002 図書館建築

1. 建築材料, 防火など
3. 設計一般
4. 書庫・書架
5. 閲覧室・研究室
7. 照明
8. 暖房・冷房・換気

例 5, 例 6 ともクラスメンバー関係以外の関係を含む。たとえば 022 を字義通り「図書館建築」と考えればその下位標数が表わす概念は何れもその種概念ではない。022 を図書館建築に関するもろもろの問題と考えればクラスメンバー関係と考えられる。

4. ISO の Guideline<sup>15)</sup>

## a) 優先関係

概念の要素分解の例には次のようにクラスメンバー関係以外の準同義関係を含む。

例 1 ANIMAL PSYCHOLOGY → ANIMALS  
+ PSYCHOLOGY

例 2 TRADE WIND → METEOROLOGY →  
+ GEOGRAPHY

すなわち例 1 の ANIMAL PSYCHOLOGY と ANIMALS の関係, 例 2 の TRADE WIND と 2 語の関係は何れもクラスメンバー関係ではない。

## b) 階層関係

クラスメンバー関係と全体部分関係の二つを認め, 両者を区別してもよいし, しなくてもよいし, あるいはクラスメンバー関係だけを階層関係とし, 全体部分関係を関連関係に入れてもよいとしている。

クラスメンバー関係の例

(上位語) 内燃機関, (下位語) ガソリンエンジン  
全体一部分関係の例

(上位語) タービン, (下位語) 羽根 (blades)

## 5. 総括

2.~4. を総括して考えると, SMART システムのような自動処理を目的とするシステムのシソーラス, UDC のような分類は事情が異なるので, その他のシソーラスについて共通して言えることは, 階層関係を原則としてクラスメンバー関係に限っているシソーラスでも, 優先関係の方は同義関係, クラスメンバー関係以外の関係を含んでいることである。

1. の例 2 で述べたようにクラスメンバー関係は同時に文献包含関係であることが多い。しかしクラスメンバー関係でも非常に上位のクラス B と下位のメンバー A との関係の場合には階層関係を付けないことが多い。これは A, B 間に文献包含関係が成立たないか, 検索上の利用効果が少ないためと思われる。

例 1. A: 放射能放出 (原子炉などから事故あるいは廃棄物処理のため放射性物質が放出される現象)

B: 放出

あるいは A, B 間には文献包含関係が成り立っても, A の下位に数レベルにわたって多数の下位語が関係付けられている場合には A, B, 間に階層関係を付けないことがある。

例 2. A: 炭化水素

B: 有機化合物

例 3. A: 動物

B: 生物

これはたとえば例 2 で言えば「有機化合物」一般の「融点」というような検索質問はひん度が低く特に上位ディスクリプタの自動付加を行なうようなシステムではこのような階層関係を付けることによりファイルが大きくなって原価的に不利だからである。

また次のような場合にも階層関係を付けないのが普通である。

例 4. A: 有機化学

B: 化学

特に, 化学を対象領域とするシソーラスの場合, 関係を付けるべきでない。何故ならば, このような場合に有機化学に関するすべての文献にディスクリプタ「有機化学」を付けることはほとんど考えられない。したがって関係を付けても意味がない。主対象分野の学問名ディスクリプタは付与する場合を限定するのが普通である。特にディスクリプタと分類を併用する場合, 主対象分野の学問名, 分野名による検索は分類にまかせた方がよい。

したがってクラスメンバー関係は階層関係付けを必要とする場合が多いが, すべての場合に必要とは限らず, 逆にクラスメンバー関係以外でも, 文献包含関係にある場合には階層付けが必要な場合が多い。しかしどのような意味的關係の場合に文献包含関係となり, 関係付けが必要になるかを研究する必要があると考える。

言いかえれば登録語間の関係を,

a) 同義関係

b) 概念的には一致しないが文献群は一致する語の関

係

c) クラスメンバー関係

d) その他の関係

に分けて考え、各々について優先関係、階層関係、関連関係の3種の参照の内のどれを付けるか、あるいは全く関係を付けないかを考えるべきである。ただし a), b) については付けるとすれば優先関係に限られる。c), d) については3種の参照が考えられ、その際一方が非ディスクリプタであるか、両方がディスクリプタであるか、文献包含関係にあるか否かが判断の基準である。

ソコロフはこれに似たような考え方で語間の関係と参照の種類の関係を表をマトリックスのような縦欄と横欄のある表で示している。<sup>16)</sup>

それではEJCシソーラスおよびUNESCOのGuideline (ISOと協同して作成)で優先関係には同義関係とクラスメンバー関係以外の関係を採用し、階層関係をクラスメンバー関係と全体一部分関係に限定しているのは何故であろうか?これはおそらく次のような理由だと思う。

(1) クラスメンバー関係がもっとも文献包含関係である可能性が強く、判定も比較的容易である。

(2) 優先関係「N USE D<sub>1</sub>」はD<sub>1</sub>で止まり、D<sub>1</sub>のBTであるD<sub>2</sub>には及ばないから、事実上はD<sub>2</sub>の範囲にNが含まれていても、作成されたシソーラス上では、たとえNとD<sub>1</sub>の関係が文献包含関係でなくても波及効果が少ないから、悪影響の生ずる可能性が少ない。たとえば、

研究の現状調査 USE 総説

の関係が文献包含関係でなくても、索引作業がこの参照通りに行われていれば、「研究の現状調査」に関する文献はディスクリプタ「総説」による探索によってもれなく検索される。ただし「総説」に関する文献の探索の際にはノイズが生ずる。

これに対して階層関係の一部に文献包含関係でない関係があると、その語対の上位の各語と下位の各語間の関係が全部文献包含関係でなくなる。

例 炭素化合物 NT 炭化鉄 NT 鋼  
NT ステンレス鋼

「鋼」は必ず「炭化鉄」を含んでいるから「鋼」に関するすべての文献は「炭化鉄」に関する文献だと考え、「炭化鉄 NT 鋼」という関係付けを行なうと、「ステンレス鋼 BT 炭素化合物」のようなまちがった関係が生ずる。すなわち「炭素化合物」

に関する文献の探索の際にノイズが生ずる。

まちがった箇所が2つも3つもあると最上位の語と最下位の語は文献包含関係から甚しく外れた関係となるおそれがある。

(3) 優先関係を同義関係とクラスメンバー関係に限定すると、すべての非ディスクリプタを同義または上位概念を表わすディスクリプタで表現しなければならないから、適合率がいちじるしく悪くなる。

例 ディスクリプタ「原子」と「散乱」があれば、上の制限がなければ非ディスクリプタ「原子散乱」を「原子」と「散乱」の組合せで表現できるが、上のような制限があれば

原子散乱 USE 散乱

としなければならない。

またこのような制限があると、「同義語」と「同義関係」のような全く同一の文献に付けられるような語を優先関係で1つのディスクリプタに統合することができなくなる。

それでは優先関係だけでなく階層関係にもクラスメンバー関係以外の文献包含関係を取入れることができないうであろうか?

文献包含関係の基準を明確にすることができれば必ずしも不可能ではないと思う。

せまい分野については比較的容易であり、MEDLARSのMeSH, EURATOM および初期のINISシソーラスではこれに近いことを実行して、かなりの成果をあげている。

理工学一般のような広い分野についても、もっと研究すれば文献の分野にかかわらず普遍的で安定した文献包含関係の基準をきめることができるとと思われる。上述のソコロフの論文では、階層関係、関連関係とされるクラスメンバー関係以外の関係を、普遍的(имманентный)関係と、暫定的(факультативный)関係に区別している。

また優先関係に限っても、文献包含関係の基準をきめることは、準同義関係の優先関係、したがって概念のディスクリプタ組合せによる表現の基準をきめるのに役立つと思われる。

6. 階層関係のその他の問題点(同一上位ディスクリプタの直近下位ディスクリプタ間の関係)

分類とシソーラスの両方を含めて一般に階層関係には語、分類記号あるいはそれらが代表する概念の対の間の関係以外に、ある語または分類記号が代表する概念と、

その直近下位の語、あるいは分類記号が代表する概念全体との関係、直近下位概念相互の関係を考慮する必要がある。

UDC などの分類についてはある分類項目の直近下位分類について一般に網羅性と排他性が要求される。

たとえば UDC では、02 図書館、図書館学の直近下位標数として

- 021 図書館の機能、効用、設立、活動、維持
- 022 図書館建築
- 023 図書館の管理、職員
- 024 利用者に対する関係、利用規則
- 025 図書館の運営事務
- 026 専門図書館、特殊図書館
- 027 一般図書館
- 028 読書、読書案内、読者心得
- 029 知的作業のための資料

があり、図書館そのものについては 026, 027, 図書館の種々の問題については 021~025, 関連する問題については 028, 029 に一応網羅されている。

ただし図書館、図書館学に関連の深い「ドキュメンテーション」、「書誌」などは別の分類項目に含まれている。したがって限定された範囲内で網羅性を持つ。

そして文献としてはこの内の複数項目に関係のあるものはあるが、個々の問題については1つの項目そして1つの項目だけに当てはまるように考慮されている。その意味で排他性を持つ。たとえば図書館は 026, 027 の何れか1つに分類されるように 026, 027 のスコープがきめられているはずである。

この網羅性、排他性を満足するため分類には「その他の○○」のような項目を設けることが多い。また複数の項目について述べた文献あるいは分類しようと思う項目の一般問題を扱った文献を分類するため「○○一般」のような項目を設けることが多い。

例 553.7 鉱水および鉱泉

- .71 アルカリ鉱泉
- .72 炭酸泉, ヨード泉
- .73 含鉄鉱泉
- .74 いおう鉱泉
- .75 石灰鉱泉
- .76 放射性鉱泉
- .77 塩性鉱泉
- .78 温泉 (定義あり)
- .79 その他: けい酸泉など

548 結晶学

- .0 結晶学一般
- .1 数理結晶学
- .2 結晶の生成・結晶の集合
- .3 結晶化学
- .4 結晶における不規則性
- .5 結晶の成因, 成長および分解

シソーラスの場合には網羅性、排他性を満足しない直近下位語 (NT1) を設けることが多い。これには2種類あって

(1) 一つの観点からの直近下位語が網羅性、排他性を満足していないもの。

(2) 複数の観点の直近下位語が共存するため排他性を満足しないもの。

(1) の例. 含有量 NT 含水量 直近下位語が1語だけの場合。含まれる物質による下分けの内「含水量」だけを採用。他にこの観点からの下分けに「含油量」などがある。したがって網羅性を満足していない。

(2) の例. EJC シソーラスの例  
clothing (被服) の直近下位語

使用目的別

- Camouflage clothing (偽装服)
- Environmental clothing (特殊環境用被服)
- Flight clothing (飛行服)
- Protective clothing (防護服)
- Tropical clothing (熱帯被服)
- Underwear (下着)

用いられる体の部分別

- Foot wear (はきもの)
- Headgear (かぶりもの)
- Hosiery (くつ下)

使用目的と用いられる体の部分別

- Goggles (防護眼鏡)

その他

- Uniforms (制服)

観点の異なる直近下位語のスコープが重複しているために、たとえば“Helmets”は使用目的別の「防護被覆」の NT であると同時に、用いられる体の部分別の「かぶりもの」の NT でもある。また‘Goggles’は用いられる体の部分別の直近下位語に「眼鏡」があればその下位語となり、同時に使用目的別の「防護被服」の下位語ともなり得る。

したがって用いられる体の部分別の直近下位語は網羅

性を満足していない。

このようにシソーラスの場合には直近下位語が網羅性、排他性を満足していない場合が多い。

網羅性を満足していない場合、すなわち直近下位語に欠落がある場合には上位のディスクリプタで代用する。

排他性を満足していないものの内、分類の観点が違うものは差支えないが、同じ観点の場合は索引作業の混乱の原因となるので、スコープが重複しないように直近下位語を変更する必要がある。また直近下位語の定義が不明確なため同一の事物に時にはA、時にはBのディスクリプタが付与されることもある。これは第二章のディスクリプタの選定条件cで述べたディスクリプタが表現する、概念の範囲の明確性が満たされていないために起ったことである。

すなわちシソーラスの場合、あるディスクリプタの同一の観点からの直近下位語のスコープの重複は絶対に避けなければならないが、異なる観点の直近上位語はどうしても必要なので、それによるスコープの重複は許容すべきである。

同一観点の直近下位語は網羅されているのが望ましいのであるが実際上不可能な場合もある。

たとえば上の例で用いられる体の部分別で「眼鏡」の他、胸部と上肢、腹部と下肢などに用いる被服に相当する語はない。使用目的別でも「熱帯被服」はあるが「寒帯被服」はない。むしろ必要なものだけを選んだという感じである。

特に個々の語の使用ひん度によってディスクリプタを選択する場合にはそのようになる。

II章のディスクリプタ選定の項で述べたようにシソーラスの使いやすさに重点をおけば同種同レベルの語（同一ディスクリプタの同一観点の直近下位語は同種同レベルと考えられる）は一括して採否をきめるべきであり、このことは網羅性を満足させることである。

## B. 関連関係

1. 関連関係の参照の数と階層関係の参照の数の関係  
ディスクリプタ間の関係の内、階層関係以外の関係を関連関係（RTなどの記号で表わす）とするシソーラスが多い。階層関係の基準にはクラスメンバー関係だけにするか、それ以外の文献包含関係を含めるかなどいろいろな方法がある。階層関係の範囲を広げれば広げる程関連関係の範囲はせまくなる。

逆にEJCシソーラスなど階層関係を原則としてクラスメンバー関係に限定したシソーラスでは関連関係の

参照が多くなる。これに対してEURATOMおよび初期のINISシソーラス、MEDLARSシステムのMeSHなどでは階層関係の範囲が広いので関連関係の参照の数は少ない。

### 2. 関連関係が必要な理由

階層関係は大部分の分類に採用されているが、関連関係は「をも見よ」参照の形で示される程度であり重視されていない。それに対してシソーラスでは1に述べたように多少の差はあるが一般に分類におけるよりも重視されているといえよう。この意味でもっともシソーラス的な関係といえることができる。

階層関係は検索の際にノイズの増加を最小限に止めながら検索もれをへらすこと、言い換えれば、適合率の低下を最小限に止めつつ再現率を高めることを目的としている。

これに対して関連関係はノイズが多少ふえても検索もれをへらすこと、言い換えれば、適合率を多少へらしても再現率を高めることを目的としている。

次に筆者が行なった金属工文学文献の検索実験の例を第1表に示す。<sup>17)</sup>

この検索実験はASM-SLA分類によるものである。以下4つの質問に対する検索式とその意味、検索された文献の数、その内該当と判断された文献の数を、関連分類項目を検索式に入れない場合（狭い検索）と入れた場合（広い検索）について示す。

この例では検索範囲の拡大により再現率が10~100%増大しているが適合率は増加しているものと減少しているものがある。

### 3. 関連関係の種類

A. の5で述べた登録語間の関係の内のdの「その他の関係」とcの「クラスメンバー関係」の一部の内のディスクリプタ間の関係であって、(2)で述べた目的に合致するものが関連関係あるいは関連参照であるが、この関係には種々のタイプがある。これらの関係を分類するには色々な考え方があるので二、三の分け方について述べる。

最初にソコロフの分類、ASM-WRUシステム analytical relation について述べ、最後に筆者の考えを述べる。

#### ①ソコロフの分類<sup>18)</sup>

この分類は関連関係に限らずシソーラス中の語間の関係一般についての分類であるが、すべての関係が関連関係になり得るのでそのまま引用する。

第1表 金属工学文献検索例

質問 番号	検 索 式 (狭)	検 索 式 (広)	検索件数		該当件数	
			狭	広	狭	広
16	08-66 (クラッドまたは接着された金属)	(08-66) + L 22 ⋮ (クラディングまたは接着)	23	33	4	8
	08 (表面処理された金属)	08 + L ⋮ (表面処理)	152	653	4	8
18	F 23 (圧 延)	F 23 + W 23 ⋮ (圧延装置)	253	302	22	24
19	L 23 (金属の溶射による被覆)	L 23 + (08-67) ⋮ (金属を溶射被覆した金属)	30	33	25	27
21	(Y Z + Z) × (09-69) ⋮ (铸铁)(鋼) (非金属介在物)	(Y I + Z + D) × (09-69) ⋮ (製鉄製鋼)	49	53	27	30

## 質問内容

16 合せ板の材質, 種類。金属材料または設計

18 分塊圧延

19 溶射被覆法

21 鉄鋼中の非金属介在物の発生原因と除去

+: 論理和, ×: 論理積,

## a) クラスメンバー関係

例 「情報処理」と「コード化」

## b) 全体一部分関係

例 「導波管」と「導波管の管壁」

## c) 機能一事物あるいは役割一操作関係

例 「確実性」と「保護」

## d) 特徴一事物あるいは性質一操作関係

例 「運動」と「動力学」

## e) 原因一結果

実例なし

## f) 類似関係

例 「重量」と「寸法」

## g) 対比関係

実例なし

## ②ASM-WRU システムの semantic code の analytical relation

これは semantic factor と semantic code の4字の構成単位との関係を示すものであるが, この両者をディスクリプタと考えればディスクリプタ間の関係を示すもので, このシステムではすべての関係を事実上階層関

係として扱っている。したがって本来方向性のある関係として選ばれたものであるが, 関連関係を分類する場合の参考になる。以下 analytical relation を示す文字と名称および実例を第2表に示す。<sup>19), 20)</sup>

## ③筆者の分類

JICST シソーラスのディスクリプタ間の該当する実例を示す。ただし JICST シソーラスには関連関係を付けていないので仮想のものである。

a) クラスメンバー関係で階層関係を付けなかったもの

例 「情報活動」と「情報サービス」

## b) 全体一部分関係

例 「言語」と「語」

c) 類似関係。同一のカテゴリー(操作, 現象, 性質, 物など)に属するディスクリプタ間の関係の内c以外の関係に該当しないもの。

## (i) 操作あるいは行為相互間

例 「索引作成」と「抄録作業」

## (ii) 現象相互間

例 「汚損」と「消耗」

シソーラスの問題点

第2表 ASM・WRU システムの Analytical Relation

文字	名称	説明と実例
A	categorical	△A△△は△-△△に対応するクラスのメンバー 例。赤は色C-L RのメンバーだからC A L R
E	intrinsic	△E△△は△-△△によって構成されるもの 例。ダイヤモンドは炭素C-R Bによって構成されるからC E R B
I	inclusive	△I△△は△-△△の部分。 例。羽は鳥B-R Dの部分だからB I R D
O	comprehensive	△O△△は△-△△に対応するクラスのいくつかのメンバーから作られたもの。 例。mosaic はいくつかの結晶C-R Sから作られているからC O R S
U	productive	△U△△は△-△△を産出するか、△-△△のために使用される。 例。ゲル化はコロイドC-L Dを産出するからC U L D。棒は打撃B-T Tに使用されるからB U T T
Q	affective	△Q△△は△-△△を利用するか、△-△△によって決定されるか、△-△△の影響を受けるもの。 例。電気めっきは電気L-C Tを利用するからL Q C T。phototropism (ある種の結晶が光によって色を変える現象) は光R-L Tの影響を受けるからR Q L T
W	instrumental	△W△△は△-△△から産出されるか、△-△△に作用するか、△-△△によって作用されるもの。 例。オリーブ油は植物B-T Nから産出されるからB W T N。焼入は金属M-T Lに作用するからM W T L。カバーは jacketed (おおわれたもの) C-C Rに作用するからC W C R
X	negative	△X△△は△-△△の不在を示す重要な特性を持つもの。 例。耐食性は、その特性として腐食D-T Rを伴わないからD X T R
Y	attributive	△Y△△は△-△△に対応する重要な特性を持つもの。 例。modified は変化 C-NGという重要な特性を持つからC Y N G
Z	simulative	△Z△△は△-△△の何等かの重要な特性を持つが、△-△△そのものではないもの。 例。鯨は魚F-S Hはないが、F-S Hでに似ているからF Z S H

- (iii) 性質あるいは状態相互間  
例 「情報需要」と「入手可能性」
  - (iv) 物相互間  
例 「図書館」と「公文書館」
  - (v) システム・学問・法則など相互間  
例 「意味論」と「語用論」
  - d) 反義関係  
例 「深層構造」と「表層構造」
  - e) 手段目的関係  
例 「典拠ファイル」と「索引作成」あるいは「情報検索」
  - f) 事物とその事物特有の性質、状態または現象  
例 「情報検索システム」と「検索効率」
  - g) 事物とその対象  
例 「質問分析」と「質問」
  - h) 原因と結果  
例 「索引作成」と「索引」(物)
  - i) 事物とそれに対する影響因子  
例 「文献寿命」と「情報需要」
- このうち理工学関係で必要とされるひん度が少ないと思われるのはc) (ii) の現象相互間の類似関係と、d) の反義関係であるが、少ないからと言って重要でないか

どうかはわからない。

金属工学文献の検索で再現率向上のため実際に使用された関連関係のうち特に多く利用されたのは上記の例のうち「操作あるいは行為相互間の類似関係」、「手段と目的」、「事物と対象」、「原因と結果」の4つの関係であった。

ソコロフの分類①との関係でみれば、a), b), d), f) はソコロフの a), b), g), d) に、c) は大体ソコロフの f) に、e) は大体ソコロフの c) に対応する。h) はソコロフの e) に類似しているが、ソコロフは e) については実例を挙げていないし必要性も認めていない。h) とソコロフの e) とは定義が違うのかも知れないが、筆者は h) は必要と考える。ソコロフの f) の類似関係は筆者の c) より範囲が限定されている。筆者の g), i) に相当するものはないが、ソコロフの c) 機能—事物, d) 特徴—事物, 性質—操作に含まれていくかも知れない。

ASM-WRU の analytical relation ②との関係では、筆者の分類 a), b), e), f), i) は②の A, O, Y, W, Q に対応する。U は e), h) に、E, I は b) に類似する。X は g) に、Z は c) に含まれる。

#### IV. 一般語の取扱い

あらゆるシソーラスについて考慮しなければならない問題として一般語の問題がある。

Ⅲ章Aの5で述べたように、既存の実用シソーラスの階層関係の多くは検索に使用しても効果のないような一般的な語を除外して、或る程度の specificity を持ったディスクリプタを最上位においている。この最上位のディスクリプタより上位の語、すなわち、より一般的な語については、種々の方法で特別な処置を行なっている。

シソーラスでは複合概念を複数のディスクリプタで表現することが多い。この場合、Ⅱ章のディスクリプタの選定条件gの処で述べたように、それらのディスクリプタとその下位概念を示すディスクリプタとの間に階層関係が付けられていなければ、下位概念を示すディスクリプタを何らかの方法でみつけ出して質問式に加えなければ完全な検索はできない。

したがって、階層関係を持たない一般語を組合せて索引作業を行っても、それらの語の論理積によってもれなく検索することは困難なので、索引作業規則でそのような一般語の組合せを禁じ、それに代る方法を採用する必要がある。

一般語の中にはこの他に特定の分野で非常に重要な少数の語がある。

この2種類の一般語を取扱う特殊な方法を以下に列挙する。

(1) 索引作業には全く使用せずディスクリプタとの関係をきめて検索のみに使用する。

例 MEDLARS システムの Non Mesh heading (Ⅲ章の2.の③の例1参照), INIS システムの検索専用語 All term.<sup>21)</sup> 後者はたとえば農業に関連のある多数のディスクリプタの表を記憶装置に記録しておいて、‘all agriculture’ という検索指令によって表にあるディスクリプタを付与された文献を全部計算機で検索するようになっている。

(2) 索引作業での使用を制限し、代りに使用する普通のディスクリプタへの参照を付ける。

例 EJC シソーラスの “Use of a more specific term is recommended: consult the terms listed below” という注の付いたディスクリプタ

(3) 少数の一般語を指定し、その用法をきめ、普通のディスクリプタとリンクして索引作業を行ない、その語対を一つの語として検索できるようにする。

例 MEDLARS システムの subheading<sup>22)</sup>

(4) 特定の語を表にして索引者が必ずその表をチェックして該当する語を付与するようにする方法。

例 MEDLARS システムの check tag<sup>23)</sup>

(5) 普通のディスクリプタのように階層関係を付けるが、注を付けて、できるだけより specific なディスクリプタを使用することを推奨する。

例 ASM シソーラス<sup>24)</sup> など

(6) 階層関係を付けず関連関係のみを付ける。

例 EJC シソーラスのディスクリプタ ‘Organic compounds’ など

以上の方法は制約の強さに種々の段階がある。たとえば、(1) は索引作業に使用しないが、(2) は使用できる。(3), (4) の subheading, check tag などは少数で、索引作業の際必ず表を参照することになっている。(5) は索引作業に使用できるが使用するかどうかは索引者の判断にまかされている。(6) は用法を特に指定していない。

もれが少ないという点では制約の強い (1), (3), (4) が有利であろう。しかし (1) は検索手段を追加する必要がある。

これらの方法のいくつかを併用することもできる。実

際に既存のシソーラスでは、EJC シソーラスは (2) と (6)、MEDLARS システムは (1) と (3) と (4) を併用している。

単独の場合、併用の場合を問わず、どのような語をどのように扱うかきめる必要がある。

たとえば EJC シソーラスでは Inorganic compounds は (2)、Organic compounds は (6) であるが、どうしてそうなっているかよくわからない。

## V. シソーラスの作成

シソーラスの作成にはいろいろの方法があるが、シソーラスはその定義 (I 章) の所で述べたように、一度作成したものを永久に使用するのではなく、たえず更新することによって真価を発揮する。作成と更新には多大の労力と出費を要する。そのことは統制されないフリータームを使用するシステムが有利であるとする議論の主たる根拠になっている。

したがって統制された自然語にもとづく索引言語を使用するシステムでも、できれば既存のシソーラスを使用することが望ましい。しかしシソーラスと索引作業規則は不即不離のものであるから、既存のシソーラスを使用する場合にはそのシソーラスの前提となる索引作業規則に従って索引作業を行なうべきである。

シソーラス作成には人手による方法と、索引作業の結果などの電子計算機による処理による機械的方法と、いろいろの段階でのその中間の方法、すなわち、半自動的方法とがある。

狭い分野の小規模なシソーラスは別として、数万語に及ぶシソーラスの作成の少なくとも語の登録、語間の関係の拡張、版下の作成には電子計算機を利用した方が有利なので、大規模なシソーラスの作成には上の半自動的方法を採用するのが現在では常識になっていると思われる。

したがってシソーラス作成法で問題になるのは主として電子計算機に入力するデータを如何にして作成するかであろう。

次に UNESCO の Guideline<sup>25)</sup> のシソーラス作成の二つの方法の部分 (5.4) を引用する。

“一般的に言ってシソーラスの作成法には二つの方法がある。この二つの方法のどちらを選ぶかはシソーラスがカバーすべき文献あるいは蓄積の型とスコープによる。これらの方法を分析的方法 (analytical method) と形式的方法 (gestalt method) とする。両者共厳密な言

語学的分析によらないと言う点で経験的 (empirical) な方法である。一般的に言って特に知識の特定領域に対しては文献の事物内容 (subject content) の分析と文献からの用語の選択を含む分析的方法が望ましい。形式的方法の場合には専門家が索引その他の参考資料のような二次的ソースから用語の候補を分析し優先すべき用語を選び用語間の関係を決定する。形式的方法はいくつかの分野を含む広はんな事物分野に対してより広く用いられるべき方法である。実際には両方の方法を組合せて用いることが多い。分析的方法はシソーラスがカバーすべき知識分野を構成する文献群の索引作業から始まる。この実験的索引作業によって得られたデータを使用して、シソーラス作成者は関連する検索システムに対して価値のある暫定的な関係を作成することができる。次に最初の文献の再索引作業と索引作業の対象の拡大によってシソーラスを修正する。

形式的方法の出発点として関連する知識領域の用語の構造を概観する必要がある。この構造とは実際上分類である。その事物内容に明るい専門家は文献を記述するため実験的に選ばれた用語を検討してその関係と構造を承認することを要求される。この場合分析的方法とは異なってシソーラスは実際の経験によってでなく、「権威者」(authorities) の貢献によって作成される”。

次にこの二つの方法をそれぞれいくつかに分けて見る。

### 分析的方法

- (1) フリータームによる索引作業結果から人手によりシソーラスを作成する方法
- (2) フリータームによる索引作業結果から電子計算機などにより自動的にシソーラスを作成する方法
- (3) 専門研究者に索引作業と同時に語の関係を作成させ、電子計算機などによりシソーラスを形成する方法  
形式的方法
- (4) 既存外国語シソーラスの翻訳
- (5) 既存の分類表、件名標目表あるいは既存のシソーラスにもとづく方法
- (6) 既存の用語集、専門家の知識をもとにし一定の方針できめた分野あるいは分類ごとに基本的な語をまず選び、それをもとにして階層関係、関連関係をきめ、specific な語を選定しながらシソーラスを作成する方法。

これらの方法をいくつか組合せた方法もある。以下 6 つの方法をやや詳細に述べる。

### A. フリータームによる索引作業の結果から人手により作成する方法

まずフリータームによる索引作業の規則をきめ、その規則によって作成した索引ファイルから一定期間の使用ひん度の統計を取り、ひん度が少な過ぎず多過ぎない語、すなわち、ひん度が一定範囲内にある語を抽出し、語間の関係を人力できめてシソーラスを作成するなどの方法である。

JICST シソーラス 1975 年版の前身である同暫定版(1973 年)は大体このような方法で作成した。ひん度は索引作業の結果を記録した磁気テープから電子計算機によって算出した。しかし索引作業の規則が不備で各索引作業者がまちまちの方法で索引作業を行なった結果を用いたので、また語間の関係付けの基準も不備だったので、このシソーラスは不完全なものであった。

### B. フリータームによる索引作業の結果から機械的に作成する方法

考えとしてはいろいろな方法が提案されているが、以下の 3 例のうち実用シソーラスに採用できる方法は 1 の①位であろう。ディスクリプタの選定と語間の関係の決定の二段階に分けて述べる。

#### 1. ディスクリプタの選定法

①一定期間の一定範囲の出現ひん度を持つフリータームをディスクリプタとする方法。上述の JICST シソーラス暫定版は大体この方法でディスクリプタを選んだが厳密に機械的に選んだ訳ではない。一定範囲の出現ひん度としたがフリータームの段階で非常にひん度の高い一般用語が除外されている場合には(多くの場合除外されている)下限だけで上限はきめなくてよい。

②一定の基準で抽出した対象文献群の抄録中のキーワードの出現ひん度と出現文献数からあるキーワードの出現総数が上記の数の文献に出現する確率を計算し、その確率と出現ひん度および出現文献数の組合せから一定の基準でディスクリプタを選出する方法。この方法は JICST で行なわれた「日本語による 標題索引技術委員会」によるドキュメンテーション関係のシソーラス作成実験<sup>26)</sup>、<sup>27)</sup> で採用した方法である。

#### 2. ディスクリプタ間の関係の決定

SMART システムの報告中の階層関係自動作成<sup>28)</sup>の方法の概要は次の通りである。まずある文献群を選び、各ディスクリプタを付与された文献の数から、次の方法でディスクリプタ対ディスクリプタの行列を作る。すなわち、その行列の要素は関係をきめようと思うディスク

リプタ対の共出現文献数を一方のディスクリプタの出現文献数で割った値である。この値は各ディスクリプタ対について 2 つづつある。たとえばディスクリプタ  $D_1$  と  $D_2$  の対には、 $D_1$  と  $D_2$  の共出現文献数を  $D_1$  の出現文献数で割った値と、 $D_2$  の出現文献数で割った値が対応する。

次に各ディスクリプタ対に対応する要素対の値を一定のパラメータ(0 から 1 までの数たとえば 0.5)と比較し次の 4 つの関係のどれかにきめる。

- |   |                  |
|---|------------------|
| a) 両要素共パラメータより小                                     | 無関係              |
| b) " " 以上   | 兄弟関係             |
| c) $D_1$ の出現ひん度を分母とする要素がパラメータ以上で $D_2$ のそれがパラメータより小 | $D_2$ が $D_1$ の親 |
| d) 上の逆  | $D_1$ が $D_2$ の親 |

第 3 表に例を示す。

C. 専門研究者に索引作業と同時に語の関係を作成させ、電子計算機などによりシソーラスを作成する方法九州大学の有川節夫氏他が開発した方法<sup>29)</sup>で、ほとんど基準を与えず研究者の主観によって、研究者が文献を読んだ時に重要なキーワードにアンダーラインすると同時にキーワード間の関係も付けさせ、その結果をそのまま電子計算機で処理して各研究者の付けた関係の総和、共通部分などを求め、同一研究者の付けた関係の総和、同じグループの研究者の付けた共通部分を求める。

前者は特定の研究者のシソーラス、後者は特定グループのシソーラスになる。電子計算機処理の際矛盾した関係、たとえば A BT B と B BT A に対しては警告が発せられあとから検討できる。

この方法は

- (i) 作成が容易である
- (ii) シソーラスの直接の利用者である研究者の知識を利用できる

と言う点ですぐれており、せまい分野のシソーラス作成には適用できるかも知れないが、広い分野のシソーラスをこの方法で作成できるかどうか疑わしい。もしできれば非常によい方法であろう。

### D. 外国語シソーラスの翻訳

一見単純に思えるこの方法もやってみるといろいろの問題がある。

#### (1) 翻訳のむずかしさ

外国語も日本語と同様同形異義語がある。原シソーラスの語間の関係あるいはスコープノートによって同形異

シソーラスの問題点

第3表 階層関係自動作成の例

文献 \ ディスクリプタ	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	
文 <sub>1</sub>	1	0	1	1	D <sub>1</sub>	$\frac{2}{3}$ ①	$\frac{1}{3}$	$\frac{2}{3}$	① = $\frac{D_1 D_2 \text{の共出現ひん度}}{D_1 \text{の出現ひん度}}$
文 <sub>2</sub>	1	1	0	1	D <sub>2</sub>	$\frac{2}{2}$ ②	$\frac{0}{2}$	$\frac{1}{2}$	② = $\frac{D_2 D_3 \text{の共出現ひん度}}{D_2 \text{の出現ひん度}}$
文 <sub>3</sub>	1	1	0	0	D <sub>3</sub>	$\frac{1}{1}$	$\frac{0}{1}$	$\frac{1}{1}$	
(文献—ディスクリプタ行列)					D <sub>4</sub>	$\frac{2}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	
1はその文献にそのディスクリプタ が付与されていることを示し、 0は付与されていないことを示す					(ディスクリプター ディスクリプタ行列)				

パラメータ 0.5 の場合

D<sub>1</sub>, D<sub>2</sub> は兄弟関係

D<sub>1</sub> は D<sub>3</sub> の親

D<sub>1</sub>, D<sub>4</sub> は兄弟

D<sub>2</sub>, D<sub>3</sub> は無関係

D<sub>2</sub>, D<sub>4</sub> は兄弟

D<sub>3</sub>, D<sub>4</sub> は兄弟

パラメータ 0.7 の場合

D<sub>1</sub> は D<sub>2</sub> の親

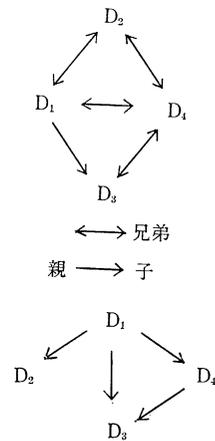
D<sub>1</sub> は D<sub>3</sub> の親

D<sub>1</sub> は D<sub>4</sub> の親

D<sub>2</sub>, D<sub>3</sub> は無関係

D<sub>2</sub>, D<sub>4</sub> は無関係

D<sub>4</sub> は D<sub>3</sub> の親



義語中のどの意味に使われているかわかる場合もあるが、わからない場合もある。また同形異義語でなくても普通の辞書ではわからない語がある。また英語では別の語であるが日本語では区別できないか区別しにくいものがある。(たとえば evaporation と transpiration など)

(2) 優先関係

外国語の同義語で日本語では対応する語が1語しかないもの。逆に外国語は1つで日本語で同義語のあるものがある。したがって優先関係のうち同義関係は語の翻訳だけではできない。

この方法で作られたシソーラスには科学技術庁がEJCシソーラスを翻訳したもの<sup>30)</sup>がある。語間の関係はTESTの磁気テープを購入し、原関係に日本語を代入して電子計算機で処理して作成した。配列は日本語の五十音順に直してある。これらのプログラムを作成して処理した。同義語で日本語が1つしかないものは同じ日

本語がディスクリプタと非ディスクリプタとなったり2つのディスクリプタになったりしている。

例 乗積 (multiplying) USE 乗積 (multiplication)

蒸発 (evaporation) ディスクリプタ

蒸発 (transpiration) 同上

このシソーラスを索引作業あるいは検索に実際に使用しているかどうか知らないが、少なくともJICSTシソーラス作成の際には参考にした。日本におけるシソーラス作成の参考資料として役立っていることはたしかであろう。

E. 既存の分類表, 件名標目表, シソーラスにもとづく方法

分類項目, 件名は「AおよびB」のような形の表現, 分類項目の場合には「○○一般」とか「その他の○○」, 「○○の○○の○○」のような表現がありディスクリプタの表現とは異なる場合が多い。それらに対しては場合

によっては分割して複数のディスクリプタに分けるか、複数のディスクリプタの組合せで表現するなどの変換が必要となる。

また階層的分類の場合には異なる階層間の関連関係を追加する必要のあることもある。

この方法で作成したシソーラスの例としては、JICSTでASM-SLA分類の分類表と、この分類による索引作業の際に作成した関連索引(逆引表)から作成した「金属工学シソーラス」<sup>31)</sup> 同様にJICSTで化学総覧の件名標目表の「を見よ」参照、「をも見よ」参照をシソーラスの語間の関係に変換して作成した「化学シソーラス」<sup>32)</sup>がある。

F. 既存の用語集、専門家の知識をもとにして選択された基本的な語の階層関係、関連関係をきめながら、シソーラスを作成する方法

EURATOMシソーラスその他ヨーロッパ共同体(EC)の作成した金属シソーラスなどは大体この方法で作成された。作成しようとするシソーラスの対象分野の専門知識ばかりでなくその分野の情報検索の経験のある専門家ががいる場合には能率的な方法である。

JICSTシソーラスはAのフリータムによる索引結果から人力によりシソーラスを作成する方法とこの方法とを併用して作成した。<sup>33)</sup>

まず問題になるのは分野あるいは分類のきめ方であるが、この分類は文献の分類でなく文献の主題あるいは事物、したがって語あるいは概念の分類であるべきである。<sup>34)</sup> その場合にもEURATOMの場合のように関連の深い語を1つの分類にまとめる方法と、カテゴリーに分ける方法とが考えられる。階層関係をなるべく同一の分類内におさめたいと考えれば、階層関係の基準によって分け方が変わってくる。たとえば階層関係をクラスメンバー関係に限ればカテゴリーに分けなければならない。実際にはこのようなシソーラスはない。EURATOMシソーラスのように階層関係をクラスメンバー以外を含む種々の関係とすればカテゴリー以外の分け方も可能である。

- 1) *Guidelines for the establishment of monolingual scientific and technical thesauri for information retrieval*. UNESCO, SC/MD/20, Paris, July 1970. 29 p.
- 2) Wersig, G. "Eine neue Definition von Thesaurus," *Nachrichten für Dokumentation*, vol. 20, no. 2, 1969. p. 53~62.
- 3) Русско-Англо-Французский Терминологический словарь по информационной теории и практике, ВИНТИ, 1968. 239 p.
- 4) *Guidelines for the establishment of monolingual scientific and technical thesauri for information retrieval, op. cit.*, p. 24.
- 5) JICST 科学技術用語シソーラス 1975 年版. 東京, 日本科学技術情報センター. 844 p.
- 6) *Thesaurus of Engineering and Scientific Terms*. Engineers Joint Council, 1967. 690 p.
- 7) Соколов, А. В. "О прагматике дескрипторного языка (по результатам анализа эмпирических микрозаурсов)," *Научно-Техническая Информация*, no. 6, 1970. p. 14~21.
- 8) *EURATOM thesaurus*. European Atomic Energy Community, 1964. (EUR 500e) 80 p.
- 9) *Medical Subject Headings-Annotated Alphabetical List*. NLM, 1976. 621 p. (PB-243 901, microfiche)
- 10) *Medical Subject Headings-Tree Structure*. NLM, 1976. 441 p. (PB-243 900, microfiche)
- 11) Lancaster, F. W. 情報検索の言語, 松村多美子訳. 東京, 日本ドキュメンテーション協会, 1976. p. 296.
- 12) Science Report. Information Storage and Retrieval. NSF [8] X-4, Fig. 1.
- 13) 国際十進分類法中間版索引. 東京, 日本ドキュメンテーション協会, 1975. 350 p.
- 14) 国際十進分類法簡略日本語版改訂第3版. 東京, 日本ドキュメンテーション協会, 1976. 293 p.
- 15) *Documentation-Guidelines for the establishment and development of monolingual thesauri*. 1st edition. 1974-08-15, ISO 2788. 13 p.
- 16) Соколов, А. В., *op. cit.*, p. 21.
- 17) 安倍浩二. "金属工学文献の機械検索に関する実験第3報," *情報管理*, vol. 6 no. 10, 1963. p. 11~16.
- 18) Соколов, А., *op. cit.*, p. 21, Table 2.
- 19) 安倍浩二. "ASM-WRU システム," *情報管理*, vol. 15, no. 6, 1972. p. 410~420.
- 20) Perry, J. W., and Kent, A. *Tools for machine literature searching*. New York, Interscience, 1958. p. 298.
- 21) Colbach, R. *Thesaurus, structure and generic posting*. IAEA-SN-128/54, 1970. p. 589.
- 22) *MEDLARS indexing manual, 2nd ed.* NLM, 1976. chap. 12.
- 23) *Ibid.*, chap. 11.
- 24) *Thesaurus of metallurgical terms*, American Society for Metals, 1968. 244 p.
- 25) *Guidelines for the establishment of monolingual scientific and technical thesauri for information retrieval op. cit.*, p. 24.
- 26) 原型シソーラス作成の研究. 索引委報告 68-1 資料 68-3, 1969年2月, 日本科学技術情報センター 日

## シソーラスの問題点

本語による標題索引技術委員会. 98 p.

- 27) 笹森勝之助. “シソーラス作成の自動化,” **情報管理**, vol. 15, no. 4, 1972. p. 260~268.
- 28) *Ibid.*, p. 264~266.
- 29) 有川節夫他. “MIR-RF 情報検索システム (I)~(VI),” 昭和 50 年度情報処理学会第 16 回大会予稿集 41~46, p. 81~92.
- 30) 科学技術用語シソーラス. 東京, 科学技術庁管理課情報室, 1972.
- 31) 安倍浩二. シソーラス作成実例Ⅱ 金属シソーラス, シソーラス作成のために, 科学技術情報活動講習会テキスト, 日本科学技術情報センター. ページD1-D10.
- 32) 笹森勝之助. シソーラス作成実例Ⅰ 化学シソーラス, シソーラス作成のために, 科学技術情報活動講習会テキスト. 日本科学技術情報センター. ページC1~C6.
- 33) 安倍浩二. “JICST 科学技術用語シソーラス 1975 年版の作成,” **情報管理**, vol. 18, no. 6, 1975. p. 463~472.
- 34) *Ibid.*, p. 467.