Library Science No. 4 1966

シソーラスの特性と利用

Characteristics and Use of Thesauri

藤川正信
Masanobu Fujikawa

Résumé

Among various techniques employed for indexing information, the approach through thesaurus organization is considered a lately developed new technique and is becoming fashion of a sort. However, the term "thesaurus" and the technique and method adopted for its compilation are not used in a consistent way. In this article, first the notion of "thesaurus" is clarified against similar products, secondly various sorts of thesauri are introduced and finally note and care to be taken for organizing a thesaurus are discussed especially in connection with mechanical or automatic processing verbal texts.

The main purpose of this article lies not in developing a scholarly discourse but in finding efficient way of compiling a thesaurus for a given objective. To carry out the task, it is necessary to pay attention to characteristics of historical development of a subject area and to the range of freedom of interpretation of a concept and an idea in the area. This means that if a thesaurus is used for linking ideas with words, it should be examined whether a standard system or classification of concepts and ideas can be usable in a subject area. If it is so, the role of a thesaurus is to provide linking between words and concepts systematically listed in a scheme and the linking may be done without too much difficulty. But if it is not so, it will be required to set up a certain number of systematic lists of concepts and ideas and each word must be identified with each concept or idea in each of the lists, thus making it much more difficult than the former.

In certain areas in natural science and technology the thesaurus approach may be very effective for information retrieval, but the same approach can be a failure in many of areas in social sciences and humanities unless enough precautions are given to a way of controlling the correlation between conceptual systems and words. (Japan Library School)

序

情報源としての文献資料の急激な増加にともない，それに対応する各種の技術的手段も急速に開発され，応用されるに至っており，特に電子計算機の情報検索への導入に沿って自動処理手法の開発研究が進められている。その中で最近注目を浴び，各所で実験が試みられ，あるいは実用化の域に達しているものにシソーラスの作成・利用が挙げられる。

シソーラスの作成に対する関心は一種の流行と受け取られるほどに高まっているが，その作成法とか利用法について充分な予備的知識を欠くために不経済な結果を招いたり，後から多大の人力作業を必要とする機械利用が行われている例が二，三にとどまらないように見受けられる。また，海外の例を見てもシソーラスの本質について異った規定がされており，その利用意図も一定していない。

Thesaurus という語が $\theta\eta\sigma\alpha\upsilon\rho\acute{o}\varsigma$ に基き，後に辞典，百科事典のごとき“知識の宝庫”を意味するに至ったことは，既に多くの人が説いている。しかし現在使用されている thesaurus は明かに辞典と区別されており，辞典が言葉の意味（定義）を与えることを主眼点とするに対し，“あるアイデアに対し，それを最も適切に表現しうる語（単・複）を発見する……”(S. R. Roget) ために作成されたものと言うことができる。

けれども文献・情報検索上のメディアとして用いられているシソーラスは，一方においてはアイデア（あるいは概念）と語（あるいは語句）との関連づけを必要としながら，他方においては文献や情報源の内容全般およびそれらの利用者群と，物理的な文献もしくはその一部あるいは処理すべきデータの集合体との関連も捕えねばならないという条件を満す必要に迫られている。すくなくとも，文献・情報検索におけるシソーラスについては，実用的有効性が大きな問題となる。

本稿においては，“シソーラス”に対する各種の解釈や意見を紹介・検討し，次にそれぞれのシソーラスの特性を吟味し，最後にその作成と利用に当っての問題点，特に機械化検索に関連する問題を考察することにしたい。

I. シソーラスの本質

シソーラスが文献・情報検索において必要と見なされる理由は，検索対象をなんらかの形で先ず明示し，次の段階で利用者が要求適合文献・情報を求める際に，先に示された検索対象が利用者の指示する概念・アイデアなどと経済的に照合されるという結果を意図するからである。このアプローチは，すでに多くの人が指摘するように，概念とそれを表現する媒体が必ずしも 1:1 の対応を示さないことに基く難点を克服するためである。

概念とその表現媒体が正確な対応を示さないという現在の状況は，もっぱら歴史的経過により生じたと見なされる。すなわち，ある概念を指示するコトバや記号は，後者が用いられた当初においては（あるいは個人と解釈してもよい）概念と正確に対応していたと考えられるが，一方では無数の概念（単純なものも，単純なものが結合した複合概念もある）に対応していちいち別のコトバや記号をその都度形成し利用することは煩雑であり不経済であるために，既存のコトバや記号を代用するという結果を生み，他方においては新たな概念が既に捕えられた概念と等しいと誤認し，したがって既存のコトバや記号を誤用したことにより，現在の混乱を生じたと見なされる。

この混乱の原因は，いづれも概念の体系の問題に結びつく。すなわち前者は概念の類似性もしくは一般・特殊に関わり，後者はそのような類形成の過程において生じた過誤である。概念は人間の知識獲得能力の発展に伴い新たに認識されるだけではなく，既存の概念の類が破壊・改変され，新たな類が形成されることにより，体系的に異った位置に属するという結果になる。また概念の類は自然に存在するのではないから，類を構成する人間の主観によっても異る。

上述のことから，次の結論が得られる。

1) 検索上の最大の難点の1つは，概念の同定 (identification) にある。同定は，人間もしくはそれと類似の思考作用形式を持つ機械の概念把握形式と，後からその概念を求める人間または機械の用いる形式の間の問題に置換される。

2) 概念の把握形式は，歴史的に変遷し，個人によっても異りうる。この問題は，概念の類の構成形式（概念の体系化）に置換しうる。

3) したがって，概念の体系が歴史的に安定しているか否か，および個人の主観が概念の体系化に当って作用しうる程度・範囲が考察の対象となる。安定している場合は，基本となるスタンダードな体系構成形式の設定が問題となり，個人差が現れる場合は，構成形式の種類の数を捕える必要が起る。

4) 個人差を生じる場合は，その差異の種類および程度により，スタンダードな形式との間に参照を付けるか，あるいは数種の概念構成形式のそれぞれの間に参照が与えられればよい。

上記の問題は検索一般に関わるものであって，特にシソーラス特有のものではない。シソーラスが問題となるのは，概念の把握あるいは指示の記号形式として日常語を用いた時だけに起るものである。また，上記の4点に見られるアプローチは，すべて“形式”という形で概念把握の統一化を図るという点で一致している。

概念とそれを指示する記号との関係は，記号の側から見れば，記号の“意味”の問題に他ならない。(“意味”については，既に本誌 No. 3 で述べたので省略する。ここでは単純に記号の指示する対象と考えておく)

あらゆる検索手段が，概念と記号の持つこの特性に関わると同時に，上記の4点からの検討を必要とすることは言うまでもない。通常の分類的手法は，不特定多数の利用者に対して，あるいは特定の利用者群に対して，スタンダードな概念の体系リストを用意し，すべての参照をそれに基いて付与するという方法である。これに対し，日常語による方法は，一方においては，件名標目表に現れているごとく，上述の如きスタンダードリストは表面には出てこないがそれを裏付けとした体系的参照を用いると同時に，コトバの使用頻度およびコトバにより捕えられる対象群の数量的関係をも考慮し，参照により同時にコントロールしようとするアプローチを取る。

分類的手法が検索効率が高いという一般的評価は，そこで用いられる検索記号のコントロールが常にスタンダードリストにより行われ，混乱が少いという事実に基くと受け取ることができる。これに対し，日常語による検索は，一方では概念の体系リストを無視し得ないと同時に，コトバの使用習慣とか上述の如き実用性が体系化と離れて問題となるために，二重性格を帯びるに至り，人間の言語使用習慣に即し，かつそれが1次的であると同時に最終的概念指示標示となる面での長所と短所を端的に露呈するに至る。

シソーラスは，この問題をコトバの側から解決しようとするものである。しかしながら，シソーラスと呼ばれるものは必ずしも統一した形式を持つものではなく，異る内容を示すことが多い。

Sharp は，シソーラスを，置換できる語（同義・類義語）をカテゴリー別に集めたものと，その逆に置換できる語の使用を禁止する意味で語をリストしたものとに二大別している。[1] 前者は Roget のシソーラスおよび American Institute of Chemical Engineers (AIChE) で代表されるように，1つの語が幾つかのカテゴリーに現れ，そのカテゴリーを中心とした‘see also’参照が与えられる形式を取る。後者は，特定の1カテゴリーに該当する語が幾つかあっても，実際の使用に当ってはそのうちの1個に限定する方法であり，Luhn の提唱した文献のエンコーディングに現れている（彼は語の代りに，コードナンバーを使用)。

このように大別されるシソーラスの本質の考慮に当って，次の事項に留意する必要があろう。

1) カテゴリーが問題となるが，カテゴリー分類表に現れる，カテゴリーもしくはファセット別の語のリストと，シソーラスの語群構成様式の比較。

2) 相互参照の種類と，そこからもたらされる結果。

3) 上述の2種のシソーラスの特性の比較。

元来カテゴリーによる語の類別化という考えは，定義を厳密にするという意図から生れたものであり，その最も代表的な形がファセット分類法である。ファセット分類においては，ある1つのカテゴリーを指示する場合にそれを単一の特性で指示するという厳格な法則が守られている。

例えば，Vickery の Soil Science に関する分類には次の如くカテゴリーが扱われている。[2]

Soil Science

The schedule which follows exhibits the following facets.

9 Kinds of soil (variously subdivided)
8 Structure (including strata, horizons, particles)
7 Constituents (chemical parent material, organisms)
6 Properties (physical, physico-chemical, biological)
5 Processes in soil
4 Operations on soil (including amendment and amendment materials)
3 Laboratory techniques
1 General

このうち 9 Kinds はきわめて厳密に次の如く規定されている。

9 Kinds, classification
b Ortho-elvium

bb Coluvium
⋮
c Primary
cb Secondary
⋮
e Intrazonal
⋮
g By origin (subdivided by a Rock or Mineral schedule)
h By climate
hb Arctic
hd Temperate
⋮
m By physiography
mb Desert
md Red
mf Gray
⋮
q By constitution
⋮

このような語の規制法を見ると，ある要求に対応するカテゴリーおよびそのサブカテゴリーにおいては，1つの語の使用だけが許されており，他の語の使用は認められていない。

このような例はFoskettの社会科学領域におけるファセット分類表を見ても同じ形式で現れており，[3] 例えばBoys やGirls はIndividual の下のPersonalities の下にだけ現れ (Be-Bi)，Son やDaughter はFamilyの下のPersonalities の下にだけ現れる。

これに対して，Roget の *Thesaurus* では，Soil をRegion, land, dirt, deface, till the—, の5つに区分し，それをカテゴリーとして見ると，region はAbstract space, land はInorganic matter の下のSpecific fluids, dirt はThe voluntary powers の下のIndividual volition の細区分の1つであるUncleanness, deface はSentient and moral powers の下のPersonal affection の細区分の1つUgliness, till the— はdirt と同じ下の細区分Precursory measures, にそれぞれ属し，そこで他の多くの同・類義語と併列的に挙げられている。

したがって，通常のシソーラスは，ある特定の語に対しそれと同・類義と見なされる語をリストすることを目的とすると言ってよい。このことをVickeryは，概念およびその関連を，同義語のコントロールおよび文章構造の単純化により規制された語の系列に変換することが検索上の問題であるという立場を明かにした上で，“Rogetの *Thesaurus* は，2つの特性——目的と形式——を持つ。その目的は，利用者がものを書く際にアイデアと語を結びつける際の助けになることである。アイデアはもちろん語によって表現されるから，*Thesaurus* はある限定された数のidea-words を遙かに数の多いtext-words と関連づけることに用いられる。この関連づけには二重の構造が見られる。第一は，idea-words が分類順にリストされる形式を持ち，その各々のidea-wordに対し同じアイデアを表現しうるすべてのtext-words が列記される。第二は，text-words とidea-words の両者が1つのアルファベット順に配列され，それにそれぞれideaを示すナンバー〔注．カテゴリーあるいはその細区分を示す番号〕が付せられるという形式を持つ。ここで，text-words が2つあるいはそれ以上のidea-words と結びつけられることがあるという事実に注目しなければならない”という意味の言葉を述べている。[4]

もしも，シソーラスをこのような意味に解するとすれば，ファセット分類法に現れる語のリストは本質的に全く異るものとなり，後者の如き性格を持つものは除外すべきであろう。シソーラスは語の選択を可能にするものであり，そこにその本質が見られる。ファセット分類法の場合，実際にあるファセットを限定するに用いられる焦点 (foci) の間の選択は許されず，どれか1つに限定する必要がある。この点で，完成された形式は類似しているとしても，前記両者は明かに区別されなければならない。

II. シソーラスの種類

前章で述べたシソーラスの本質を，各種の形式とか程度により備えているものがあるが，それを大別すると，次の如く分類できる。

1) 語，句の間の参照 (see, see also) が与えられており，概念と語との関係は参照を通じて規制されているが，語が示す概念の距離は与えられていないもの。自動翻訳に用いられるシソーラスはこの類に属する。

2) 同義語，類義語間の規制が，関係用語を直接関連づけ，またその語句が示す概念の上下，大小などを表示する表現が取られ，かつそれらの語句で代表される概念の相対的位置が別に設けられた分類表の記号で示されているもの。通常の件名標目表がこの類に属する。

3) 上と同様であるが，概念の体系的関係が語句で示されているもの。D.D.C. (ASTIA), Eengineers Joint Council (E.J.C.), Bureau of Shijps (B.S.) などのシ

ソーラスがこれに属する。

4) 2) とほぼ同様であるが，語句のリスティングにカテゴリーを持ちこみ，概念の相対的位置をそのカテゴリー番号で示す。MeSH (N.L.M.) の方式はこれに属する。

5) 語句の使用上の関連を別の語句で指示することは全く行わず，別に設けた体系記号によって処理する。この処理は計算機により自動的に行われ，一部の固有名詞およびそれに類するものを除き，処理結果はすべて上記体系記号で表示される。わが国の外務省の方式はこれに属する。

文献・情報の検索に当って利用されるシソーラスは，このうち 2)～5) のタイプに属するものであると見なされる。1) は通常の同義語辞典と本質的に同じであり，2) 以下の作成に当って参考とはなるが，検索に直接用いるには，概念の位置的関連が捕えられないので不便である。したがって，2)～5) を具体例により検討し，その利害得失を論ずる。

2) に属する件名標目表をシソーラスと見なすか否かは，シソーラスに対する定義をどこまで厳密に適用するか否かにより決定される。Descriptor と称せられるものが，実際は件名標目と異らないことが発見されると同様に，シソーラスと称して件名標目表と内容的に同じであるものも少くない。また，件名標目表に用いられている技術が，シソーラスの作成に当って利用される面もあるので，いちおうここでは取り上げることにする。

件名標目表における 'see' 参照は，かくべつの問題を提起しない。問題となるのは，'see also' 参照である。先に挙げた Soil の例をとり，Sears の表に徴すると，次の如く表現されている。

Soils

see also **Agricultural chemistry; Alkali lands; Clay; . . . Reclamation of lands; Soils (Engineering);** also headings beginning with the word **Soil**

この指示に基いて **Clay** を見ると，次の如くになる。

Clay

see also **Bricks; Modeling**

Bricks

see also **Bricklaying; Tiles**

このような追跡形式をどこで止めるかは，索引を利用する場合と同様な考慮を要する。索引については，*Applied science and technology index* の **Diesel engines** につき同様な追跡を行ったところ，200 近い検索項目が得られた。このような追跡は実は無意味なものであるが，検索語が相互参照形式で与えられていて，自分の求める概念やアイデアを示す語が見出せない場合は，それを発見するまで追求を終らないという態度を取ることもまた自然であろう。

無意味な追跡を繰り返さなければならない理由は，検索用の語句が示している概念が，概念の一定のシステムの中においてどの位置を示しているかを発見できないからである。**Soils** を土壌科学，農業，不動産，工学などのどの立場かで捕え，その上位，下位の概念を与えておけばこの問題は防止できる。'See also' の形の参照を増すことは，利用者に対して決して有利な方法とはならない。むしろこの形式の参照をできるだけ減少し，逆に 'see' の参照を概念を中心として設けるほうが有効度は遙かに高くなると解される。その極端な形が，5) の外務省の形式に現れる。

同時に，シソーラスにおいて，そこから求められる語句を coordinate indexing 方式に基いて結合して利用するという原則に立つと，論理積の関係が非常に複雑となり，特に機械検索の効果を甚しく低下せしめる結果を招くことになる。

件名標目表に現れている語句の規制方式は，一般の図書館などで用いる場合は，利用者が慣用する語句の多様性に応じうるものとして利用価値を持っている。しかし特定の専門領域で概念体系が問題となる場合は望ましい形式ではない。

D.D.C. (ASTIA) のシソーラスは，これまでにも各方面で紹介されてきたが，その正式の名称は *The thesaurus of ASTIA descriptors* で第 2 版が1962年に発行されている。このシソーラスにおける Group および Field Structure は，システムと利用者のそれぞれの扱う語彙の調整のために設けられたもので，その機能は Roget の categories に相当する。ただしその構成はかなり異っており，D.D.C. の場合は Detection, Defence, Abstract Concepts および数学，力学，航空工学などの学問領域を指示するカテゴリーに分けられている。

またこのシソーラスにおける 'Use' reference は，システム語彙に利用者の語彙を集約するために設けられた。さらに，各語に対しては定義と Scope note が必要に応じて与えられているが，その例は B. S. のシソーラスにおいても見られる。

これらすべての特質は，次の 3 点にまとめられる。す

なわち：

(1) 語の間の関係を表示する

(2) 必要な定義を与える

(3) コンピューターに対するインプットの際のコードブックとして用いるの

(1) の関係は，同(類)義：体系：一般の3者を扱い，それぞれ次に示すごとき扱いをしている。

同義： Acceleration integrators
Use Accelerometers
Electric code
Use Electric wire

Acaricides
(Pest control and inhibiting agents)
Includes Miticides
Barometric pressure
Includes Atomospheric pressure

'Includes' の指示は，用語上もはや使用を停止した語，句を指示する場合にも用いる。

Nuclear reactors
Includes:
Atomic piles
Atomic reactors
Standing-wave ratios
Includes:
SWR

体系： Blood plasma
specific to plasma
Plasma
Generic to Blood plasma; Gaseous plasma.

一般： Accelerators
(Particle accelerators)
Also see Betatrons
Cyclotrons
Electron accelerators
⋮
Synchrotrons

これまでに例として挙げたものに AIChE の *Chemical engineering thesaurus* (C.E.T.) を加え，それぞれに用いられている用語規制形式を比較すると，次の結果が得られる。[5]

参照形式として見た場合の最も大きな特質は，従来の件名標目表で一般的に使用される例にくらべて See also の内容が下の4例においては詳細に区分されている点である。類義語の関係を，たとえ別の言葉に置き換えるに止まったとしても，その操作に体系的要素が明示されるならば，第Ⅰ章で述べたような無益な See also をたどる追跡は防止できる。

4) の MeSH は，上例と異り，カテゴリーとその細分の中における語の順序により，体系的関連が捕えられる。その構成は通常のカテゴリー分類表における各項目の配列となんら変りがなく，組合せの様式も同様である。

カテゴリーはA—Mの13種であり，このうちカテゴリー A (Anatomic Terms), B (Organisms) および C (Diseases) に関しては特に詳細な使用説明が加えられ，カテゴリー D (Chemicals and Drugs) までの項目数総計は全体の約 $\frac{3}{4}$ を占める。

各カテゴリーは，F (Psychiatry and Psychology), H (Physical Sciences), I (Social Sciences), J (Tech-

図書館件名使用	D D C	B S	E J C	AICHE
See	Use	Use	Use	See
See (x ref.)	Includes	Includes	Use for (U. F.)	See from (S. F.)
See also (xx ref.)	Specific to	Broader terms	Broader terms (B. T.)	Post on (P. O.)
See also	Generic to	Narrower terms	Narrower terms (N. T.)	Generic to (G. T.)
See also	Also see	Related terms	Related terms	Related terms

nology, Commerce and Industry), K (Humanities), L (Communication, Library Science and Documentation) および M (Named Groups of People) を除き，サブ・カテゴリーに細分され，それぞれを構成するトピックが列記されている。

例： A- ANATOMICAL TERMS
A-1 Parts of the Body
ABDOMEN
BACK
EXTREMITIES (A 2)
HAIR
HEAD
NAILS
NECK
PELVIS
SKIN
THORAX (A 2, A 4)

上例のうち (A 2) を付せられた EXTREMITIES は，A 2-Musculoskeletal System においては，(A 1) が付せられており，この項目名はカテゴリーの取りかたにより，A1 にも A2 にも属することを現わしている。同様なカテゴリーあるいはサブ・カテゴリー指示記号が，該当項目にはすべて付記され，参照もしくは，カテゴリー決定上のキーの役目を果している。

共通形式区分としては abstracts, indexes, statistics など24項目が用いられ，別に閲覧用目録中の同様な形式区分（施設名などを含む）は地域名あるいは国語で細区分されている。

項目名間の相互参照は次の形式をとる。

他の語，句への参照	他の語，句からの参照
See	X
See under	X U
See also related	X R
See also specific	X S

このうち See also related は広く関連項目を利用者に示唆する役目を果し，See also specific は generic→specific の指示に使用される。

例： SCIENCE (G 1, H, I)
See also related
RELIGION AND SCIENCE (K)
RICINUS (B 6)
See also specific
CASTOR OIL (D 6, D 8)

実際の使用は，次の如き形式を取る。

(1) 探索項目

"Diseases of the nervous system" in association with "protein metabolism and diseases of the nervous system" especially in relation to "the aminoacidurias" and particulary "phenylketonuria."

(2) MeSH 項目

C 1 〔diseases of the nervous system〕
〃 2 〔protein metabolism disorders〕
〃 3 〔albinism〕
〃 4 〔albuminuria〕
〃 5 〔amino-aciduria〕
〃 6 〔alkaptonuria〕
〃 7 〔cystninuria〕
〃 8 〔maple syrup urine disease〕
〃 9 〔phenylketonuria〕
〃 10 〔amyloidosis〕
〃 11 〔cystinosis〕
〃 12 〔ochronosis〕

これらは，それぞれCのカテゴリー中のサブ・カテゴリーから集めたものである。これらの項目は質問内容に応じて，論理的関係を構成する。(1～10は仮り順序をつけるための番号にすぎない)

C1 ∧ (C2 ∨ C3 ∨ C4) ∧ 〔(C5 ∨ C6 ∨ C7 ∨ C8) ∨ (C9 ∨ C10 ∨ C11 ∨ C12)〕[6]

このうち〔 〕内は，すべて ∨ (論理和) で結合しても変りはない。

MeSH 方式の最大の特徴は，カテゴリーおよびサブ・カテゴリーの段階までは分類形式をとり，それ以下の項目配列はＡＢＣ順となり，各項目はさらに細分が必要な時は再びＡＢＣ順に細目を記載し，探引構成の際はカテゴリーを問題にはするが実際には探索概念に対応する語を集め，それを論理的に結合する方法に見られる。

5) の外務省方式は，自動処理のために作られ，体系的処理は調書管理コード表を基本とし，漢字仮名まじりの語・句から成り，非体系語を含みうるという点に特質を持つ。

自動処理は，索引語・句に関し，それがすでにシソーラス中に含まれているか否かの自動認識と，含まれている場合はその語句を体系コード化する作業（体系語に限られる）と，簡単な文法チェックに関し行われる。現段階では索引語・句の選び出しは，一定の索引形式に沿っ

て，人間の手で行われる。将来は，その自動的な付与が意図されている。

調書管理コード表は，主体をなすのが「主題記号」であり，その他「地域・国名記号」「局課・在外公館記号」「役割記号 (Role indicator)」「年月日指示記号」などから成る。シソーラスにおける体系的処理は，このコード表を改訂増補した結果と見てよい。

シソーラス収録語彙は，漢字・仮名の混用語・句であり，入出力に漢字テレタイプを使用する。この形式を採用した理由は，母国語を使用できる便利さ，ローマ字化した場合の同音異義および長文化の問題に対処するためである。

上記の特質が，体系語の処理に際して，語・句を直接コードに転換するという形式を取った理由になる。

例：
マラウイ →※ A 43830
Malawi →※ A 43830
マラヤ →※ A 12151
Malaya →※ A 12151
マラヤ連邦 →※ A 12151
祖国防衛
→ (そ国防衛) ∪*B4000000
∴※ B 3800000 (安全保障)
※ B 3900000 (軍備縮少・軍備管理)
※ B 3A00000 (紛争・戦争)
↓ +B4#

(∴ は see also, ↓ は specific to と見なせばよい)

索引形式は，主体，客体 (ただし，形式的な区分としては，同一と見なす)，内容 (主題)，場，時，出所，形式 (発表，文書) などのほか能動，受動，その他の論理関係などがそれぞれカテゴリーを形成し，内容的まとまりを個々の clause と見なす。シソーラスの語・句はこれらのカテゴリーに1つ，またはそれ以上現れることになる。[7]

以上，シソーラスを第 1) 形式を除き4型に区分し，簡単にその特性を記したが，2)-4) 型と 5) の外務省は構造においても，機能においても全く異質な点を持ち，それが利用および索引効率の面に大きな影響を与えることに注目しなければならない。

III. シソーラスを用いる検索形式

外務省の例を除く場合，2)～4) 型はすべて語・句自体に体系記号が付けられておらず，MeSH の場合といえども，付記されているカテゴリーまたはサブ・カテゴリー記号は参照のためのもの，もしくはカテゴリー選択の媒体にしか過ぎない。

そうすると，索引の作成に当っては文献のテキスト中の語の中から適当な語・句があればそれを抽出し，もし無ければ補足して語・句を集めるが，その場合の集成形式は基本的には論理積となり，論理和は積の内部，もしくは中間においてのみ現れる。この場合の論理形式は，当然テキストの文脈に沿って形成される。

次に，要求に適合する索引形式の作成は，情報要求者の質問の文脈に沿って語・句を選び出し，それを上記同様に基本的には論理積の形式にまとめる。

この結果は，coordinate indexing の形式となんら変りのないことは，第Ⅰ章においても問題提出という形で触れておいたとおりである。Coordinate indexing においても，単位概念 (unit concept) を表示する単位語 (unit-term) はリストとして作られており，同義語および類義語の関連，およびある程度の体系的関連は語・句の間で付けられていると見なければならない。

すなわち，coordinate indexing 方式は，検索操作において単位語を論理積の形で処理する (当然原文献の索引作成に対応する) 索引形式が問題であって，単位語の集成されたリストの形式を問題とするものではないが故に，語・句のリストの形式から焦点を当てれば，それがシソーラスと呼ばれるものであっても支障はないということになる。

シソーラスの側から見ると，リストの中の語・句に各種の参照が付せられていようと，それらの参照は文献に対する索引語の抽出および質問に応ずる際の検索語の抽出の際に利用しうるだけであって，それらの参照は選び出された語・句の結合の際には，せいぜい論理和として併記される語・句を選び出しやすいという点だけが有用であり，したがって検索操作は coordinate indexing 方式となんら変りがないと言わざるを得ない。

以上の理由により，シソーラスの利用による検索形式は，外務省の例を除き，通常は coordinate indexing の検討により得られると考えてよい (外務省の例の場合は後述する)。

Coordinate indexing の最大の問題は，Sharp の論じている次の例に最も端的に現れている。[8]

もしも，航空力学に関する語が，次の如くリストされ，縦の列がそれぞれ関連の深い語もしくは類義語であるとし，かつ下記の質問が提出され，それに対応する語を選ぶとその結果は付記した馬鹿げたものとなる。

(1) 索引用語

Wings	Delta	Interference	Speed	Scoops	Calculation
Aerofoils	Triangular		Mach number	Inlets	Estimation
Lifting surfaces	Dart-shaped			Intakes	Determination

(2) 質　問

"The estimation" of "the effect" of "Mach number" on "the interference" between "scoops" and "triangular" "aerofoils". (" " 筆者加筆)

(3) 検索用語の組合せ

$3\times3\times1\times2\times3\times3=162$

これは Sharp の言うように，極端な例であり，このようなことを防ぐためにシソーラスが作成されると見なければならない。しかし，原理的には coordinate indexing 方式にはこの危険が潜んでおり，シソーラスの作成においても充分注意を払わなければならない。

上記の例に関しては，2 つの問題を抽き出すことができる。その 1 つは redundancy に関わり，残りの 1 つは word association の形式に関わる。

Redundancy を意識的にシソーラスによる検索方式に採用している顕著な例は既述の AIChE の場合であり，検索に当って検索担当者が最も適切であると思った語・句に付せられている R.T. の中から該当する (relevant) であると思うものを選び，付加的に使用する点に見られる。索引作成の場合にあっては，R.T. は適切であると思った語がそうであるか否かを，他の参照事項と併せて，判定するに用いると考えてよい。索引作成作業と検索作業は，等質と異質の点を含むのであるから，区別しなければならない。Bar-Hillel は，この問題を coordinate indexing の理論的基礎に関し批判しており，redundancy は検索作業で主として重要な役割を演ずることを明示している。

Redundancy は言語学的に見れば Semanteme の処理に関わると見られる。Semantemes の制御については，Bernier と Heumann が次のような意見を述べている。[10]

Semantemes 間の関係を情報検索面から捕えれば，比較的少数の種類に限定できる。

(a) Synonym, (b) antonym, (c) group, genus or class, (d) kind, species, or subclass, (e) part of a whole, (f) a whole made up of parts, (g) mathematical function, (h) multiple relationship.

このうち (b) は通常のシソーラスでは取り扱わないですむ。(c) は大 (高次) 概念，(d) は小 (低次) 概念，(e) は部分，(f) は全体，(g) は関数と考えればよく，(h) は (a)～(g) のいづれかまたはその組合せで表現可能となる。したがって，それぞれを次のように記号化することが可能である。

(a) =, (b) 1/, (c) >, (d) <, (e) }, (f) {, (g) f.

この他に"関係"として考慮する必要があるものに，科学などにおいて未だ明確な関係の規定が行われ得ないものが挙げられる。この場合は，関係をつけないという形式を取らざるを得ない。次に，morphemes の問題があるが，これに関しては種々な検討を必要とするであろう。Bernier などはこれを無視しているが，KWIC 方式の利点は，morphemes を通して文脈構成要素を捕えうるということに発見される。すなわち，彼らが問題にしている "reactions in benzene" と "reactions of benzene" の区別ができるという結果を生む。

この点に関しては，Farradane の意見が参考になるが，[11] 彼の提唱する分類法は，第一に類概念を語の連合によって全く帰納的に作り上げることを意図し，演繹的手法を無視している点で問題があり，第二にアイデアとしては理解できても semantemes 間の関係を具体的に捕えていない所に難点がある。

Morphemes が原テキストの形で，制限字数以内で扱われる例は，KWIC 方式に見られるが，morphemes により文脈が捕えられるという点では，KWIC 方式は coordinate indexing 方式よりもはるかにすぐれていると考えられる。

Black は，keyword に関する考察の中で，"体系的細分の論理性は，実際には存在しない場合が多い。多くの学問領域に，技術が重複して現れるという事態が，現在の〔検索〕システムに大きな圧力を加えている" と述べた後で，"ある主題を記述する語または語の集合は，それぞれの分野における究研者が実際に使用するものである。これは，もちろん，Luhn が 'Keyword' または significant word と呼ぶものである" と見なし，[12] KWIC 方式の利点を説いている。

Black は，そこから concordance との関連を説いた上で，"不完全なタイトルを補足する一方法は，〔内容に即する〕要語をタイトルの後に付加し，括弧でくくるこ

とである”とし，タイトルに現れる語の他に descriptors を設けておいて，その中から要語を選んで付加するとすれば，最高の検索効率を得ると思われると述べ，次に従来の体系分類を用いた場合のコストと検索率を比較し，次の表を掲げている。[13]

	伝統的方法（例：UDC）	KWIC方式（自家計算機使用）	KWIC方式（データセンター利用）
2,000 文献処理費用	£ 500	£ 200	£ 350
効率	82%	76%	76%

さらに彼は文献語（テキストに現れる語）と利用者使用語の関係をつけるためのシソーラスの必要をSDI方式について触れているが，[14] これを拡張すれば，各種の同義語の関連をつけねばならないことになり，KWIC方式の持つ機械的簡便性が消滅し，シソーラス方式に吸収される傾向を示すことになる。

Semantemes と morphemes の両者を索引および検索方式について考慮すれば，結局は先述の word association の問題に帰すると考えてよい。

Word association を制御する方法には2つ考えられるが，その1つは morphemes を残す（前置詞，接続詞，語尾変化など）方法であり，他の1つはカテゴリーの枠を設け，前置詞の役割を代行させる方法である。後者はわが国の外務省の取る形式であるが，これについて論じているものに Papier と Cortelyou がある。[15]

彼らは，カテゴリーを fields, objects, processes, properties, substances の5つに区分し，それにより同・類義語および体系的処理などを可能にしている。同時にそのような処理を経て作成されたシソーラスにおいては，シソーラスの別の部分に現れる関連語（R.T.）の association list を付加するが，これは AICHE の R.T. および従来の件名標目表に見られる‘see also’と同性格のものである。

この処理操作においては，利用者が思いつきにくいと思われる関連語の把握を可能にする注意が払われている。

実験においては，各領域の専門家を選んで，任意の順序に配列された一定数の術語に対し，association list の中から第一に選んだ語を抽出するという形で行われた。この際に見られた特色は，一般的な語に対し，association list の中から特殊性の強い語（glycol→aldehydes, chemistry→chromatography）が選ばれた点に見られる。

結果から見ると，Properties のカテゴリーを除き，R.T. が同一カテゴリーから選ばれる確率が高く，また Process と Substance のカテゴリーの相互関連性が大であることが判明した。

この結果は，word association においてはカテゴリーその他体系的な処理が行われれば，索引語あるいは検索語の選び出しの分散度が小さくなり，従って両者の関連も，そのような考慮が払われていない場合にくらべれば，当然緊密になると予想される。

通常の coordinate indexing においては，morphemes も保存されず，そこで用いられる unit-term に関し体系的処理および／あるいは関連語の処理が予め行われてそれを使用しない限りは，単なる論理積の関連だけで概念の結合が行われ，語の使用から見れば，当然ノイズを生ずると受け取れる。

これに対処する方式を論じたものに，Jolley の発表がある。[16]

通常の aspect 方式，feature card 方式あるいは look-up 方式はマトリクス形式として見れば，次の形で図示できる。

	0	1	2	3	4
A			●		●
B	●	●		●	
C					●
D			●	●	
E	●	●			

上図において，0, 1, 2……は文献番号を，A, B, C……は Unit-term または feature を表わすものとする。各文献に feature が2つづつしかないというのは例外的であるかもしれないが，判別を容易にするためと理解していただきたい。（原著者注）

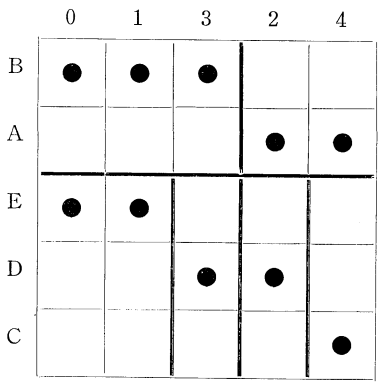
Jolley の思考法は，最後においていささか混乱しているように見受けられるが，上記のマトリクスを捕える際に，items に対して entities を，feature に対して attributes を置き換えて考える点に興味が持てる。この

考えを coordinate indexing 方式に適用するに当って，Jolley は orthophores と diaphores というアイデアを持ちこんでいるが，この両者はそれが同時に存在することによってマトリクスを形成すると受け取ればよい。いま orthophores を文献を意味するものとし，diaphores を unit-concept または features を意味するものとすれば，この両者が重要であるのは，“われわれの言語は features を持つ items については記述の用に供せられるが，逆に items を所有する features については触れない，つまり Bill は好奇心を持つ，とは言うが，好奇心は Bill という人間を持つ，とは言わないという特性に甚く偏った思考習慣の持つ危険性を取除く”[17] 点にある。

次に彼の思考法で注意すべき点は，diaphores が orthophores と交ることによって，前者が後者に2つの特性すなわち同定 (identification) と比率 (proportion) を付与するという見方である。“同定は，どの diaphores と orthophore がプラスの形で（便宜的には，マイナスの形で考えてもよい）交るかという問題であり，比率は，orthophore が幾つの diaphores とプラスの形で交るかによって捕えられる〔プラスおよびマイナス形というのはビットの有，無と受け取れる。筆者注〕…… 同定と比率はマトリクスのそれぞれ1面を代表しており，前者は論理に，後者は統計に本質的に関わり合う……これが，coordinate indexing 方式の半面がドキュメンテーション活動の埒外にあると見られる理由である”[18] と Jolley は述べている。

同定の問題はまさに coordinate indexing の問題であり，ある特定の feature(s) を持つ item の探索に用いられるが，比率の問題は indexing に関わるものではなく，item cards と feature cards の集合の総体における両者の相互関係において現れると理解したほうがよい。もし indexing の問題と関連させるとすれば，上述の関係を indexing system の検討とか改善に反映させるという形を取るであろう。

第三に Jolley が導入している考え方は，analytical system である。このシステムは，ある item が持つ feature のうち最も重要なものを選び，重要なすべての features を相互排他的に，しかも集合的には網羅する形で配列し，その順位に応ずる記号を付けることを基本とする。次に，第二に重要な features に関し同じ手順を繰り返す。例えば 597 という記号を付せられた item は，最も重要な features のグループの第5番目と，次に重要なもののうち第9番目と，三番目に重要なもののうち第7番目を同時に所有することを示す。これはしかし，ある item に分類番号を付与することとなんら変りはなく，coodinate indexing とは全く別の検索手法を突然持ちこんだことになり，本質を逸脱した議論となる。したがって彼がその後に掲げて説明しているマトリクスの形式変換は，格別の意味を持たなくなる。[19]



上図において，太線は領域区分を示す。AおよびDの features を有する新しい item は2と4の間に位する（原著者）

このような再配列は，やはりその結果を検索タームのシステムおよび文献の検索深度の測定などには利用できるが，この検索方式そのものを問題にすることではない。

以上，coordinate indexing 方式が持つ長所と短所の理解およびその解決に関する幾つかの意見を検討してきたが，シソーラスと関連させて考えると，以下のようにまとめることができる。

1) シソーラスは，概念の体系的配列とは本質的に異る。

2) 語，句が示す (meaning, reference) アイデアあるいは概念の体系を無視すると，意味上の混乱が生じる。

3) この混乱を防ぐためには semantemes の処理に当って，意味と概念体系の両側面からの規制が必要である。

4) 各語，句は単独の形でシソーラスに収録されるとしても，索引および検索作業に当ってはそれらの結合方式を考慮しなければならない。

5) 結合は，原テキストの文脈を保存する形式であると同時に，索検時に当って原形式に適合しやすい考慮を払う必要がある。

6) 結合様式は，原テキストの側に重点を置けばKWICあるいはその改訂方式に近いものとなり，索引のマッチングに重点を置き，かつ文脈形式に一定性を保たせるには，外務省が利用しているカテゴリーの一定順序による配列様式に長所が認められる。

7) 新事実，理論の展開による新概念に対応するためには，シソーラスには常に改訂の必要が認められ，そのための準備作業を継続的に行う必要がある。

8) 新事実，新概念の導入は，体系分類において生じうる根本的変動はもたらさないが，意味上の連関については処理しなければならない。

9) 語，句の記号化は，日常語形式を基本とし，語，句の使用習慣および意味上の使用規制は参照を付与することで解決される場合が多いが，検索時に当って体系的に上位下位の概念を一定限度内で連合させるには，それに応じうる体系記号の利用が有利となる。

10) “参照”は同義語間においては問題を生じないが，類義語間においては次の考慮が必要となる。

(a) 類義語の収集に当っては，文脈中の意味が問題となるから，少くとも第一段階においては，文脈を保存する形式で材料となるものを記録してそれを加工することが望ましい。

例：X国の軍事情報収集機構（情報—軍事）
Y国諜報機関の活動現況（諜報—政治）
産業情報の収集と処理（情報—産業）
The craft of intelligence
（インテリジェンス—政治，軍事）
電子計算機によるインテリジェンス処理
（インテリジェンス—電子計算機）
企業における情報と決定行為（情報—企業／決定行為）
人間の組織とコミュニケーション
（コミュニケーション—組織）
マス・コミと読書
（マス・コミュニケーション—読書）
テレ・コミュニケーションの最近の発達
（テレ・コミュニケーション—通信工学）
⋮

右側の（ ）内は，それを集めた上でカテゴリーにより処理すればよい。

(b) 類義語の範囲は，検索時に当って必要なものをもれなく集められると同時に，不要なものを用いないことにより検索効率を高めうるように定めなければならない。そのためには，カテゴリーを区別しうる指示記号を付与しておけばよい。その記号は日常語であってもさしつかえない。

例：情報

諜報	（軍事，経済，政治）
インテリジェンス	（同上；計算機工学）
コミュニケーション	（通信工学，社会学）
⋮	

“情報”そのものについては，次のような処理を行うことができる。（一種の R.T. と考えてもよい）

情報

蓄積・検索	7
評価	1
研究・開発	13
機構（システム）	8
処理	8
担当者	2
⋮	

（右側の数字は，左側の事項を扱っている文献数。これにより，検索時において利用者は自己の要求する資料の有無を知りうる。ただし，この数は機械の利用などにより常に改正されねばならない）

IV. 検索機械化とシソーラス

電子計算機の利用が文献情報処理において著しく発達すると共に，シソーラスの利用も増大してきている。

シソーラスの利用を具体的に理解するために，Swanson が挙げている例をとる。[20]

(nuclear power / or nuclear energy / or reactor / etc.....) and at the same time contain

(commercial practicality / or cost per kilowatt / or commercial profitability / or economic feasibility / or private enterprise / etc.....)

上記の形式は，提示された質問を論理関係を入れて整理したものであり，中間の句が左，右の語，句を論理積の形で結合している。ここで注意してよいのは，Swanson も述べているように，文法形式は問題になっていないということである。その理由は，文法形式の認識はテキストの探索に当って要求されていないからであると彼は述べている。[21]

この方式に従えば，質問をインプットすると，記憶されたシソーラスまたは辞書の中から必要な語句を見出

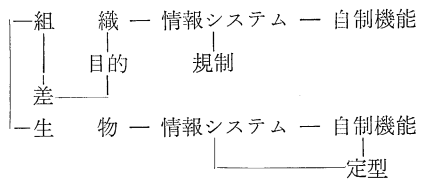
す。辞書中の語句には，人間が予め用意した weight が付与されているか，あるいは必要な参照または体系記号がつけられている。Weight は検索に当って，その語，句の重要度を判定するためのものである。

ここで問題になるのは，上記の例のごとく (A ∨ B ∨ C……) ∧ (α ∨ β ∨ γ……) という単純な形で質問が常に処理されるか否かということである。( ) 内の項目の順序の変更はこの場合大きな問題にならない。

いま質問として,「人間の作る各種の組織において，情報システムが自制機能を持たないかぎりは，規制機能をなんらかの目的にそって作り上げなければならない。そのような目的と規制方式は,組織の差を反映するものか，あるいは組織が異っても規制機能は異らないものか。異らないとした場合は，生物体に見られる情報の規制のように，ある定型を捕えられるか」というアイデアから出発して項目の関連を形成してみるとする。

先ず項目を独立した形で選び出すと，組織，情報システム，自制機能，規制 (制御)，目的，差，／生物，情報，規制／定型，となる。

これらの間に関係を設けると，



この図式中で，——はすべて論理積の形をとり，それぞれの語，句に対し類義語は論理和の形をとる。もしも論理積の形式が横に線型で捕えられるならば問題は少ないが，二次元的形式を取る場合は，それに別な操作を加えて一次元的に変換する必要がある。

| 組織 ∧ (目的 ∨ 差，目的 ∧ 差) | 情報システム ∧ (規制 ∨ 自制機能)

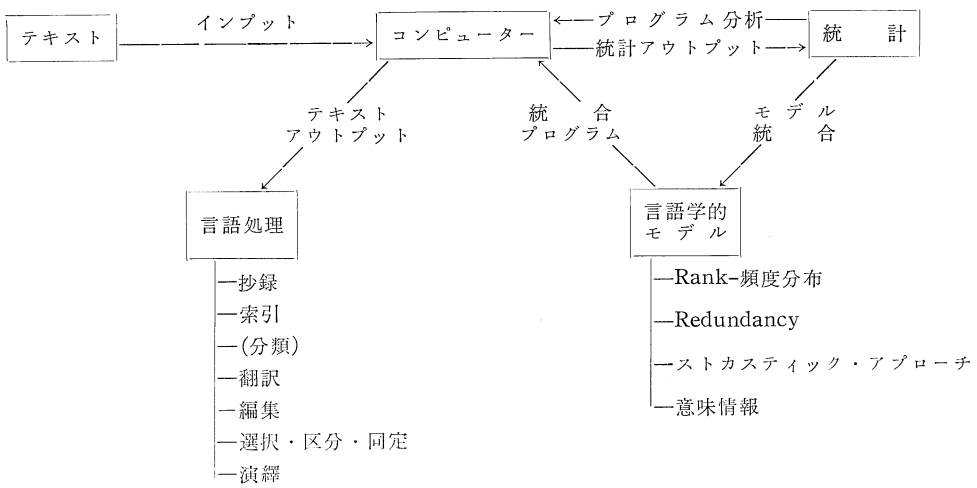
このように区切ると，field または section という考えを入れ，その中であるアイデアに関する事項が1ヶ所にまとまるように考えたほうが有利であろう。

また，生物に関する事項は別のまとまりであり，検索に当っては，第一の情報に関わる部分の検索が終った後で，その結果の中から (テキスト判別をした上で)「規制機能は異らない」というものについてのみ，生物の情報システムの比較をしているものを選び出せばよいことになる。したがって，この質問の場合は，検索に当っては情報——と生物——とを link させる必要はない。Link の必要が生じるのは，索引作成の際のみである。

シソーラスは索引，検索のインストラクションに用いられるものではなく,索引・検索に当って使用される語，句の選出に使用されることは言うまでもない。しかし，情報検索全般から見た機械利用を考慮し，全機能が有効に発揮できる考慮はしておく必要がある。

このような機械による言語情報処理システムを予想した場合，シソーラスは各種の個別処理単位に関連していることが分る。したがって，機械の利用を考えれば，このような処理単位で必要とされる各種の操作上のキーを記号化してシソーラスに含ませ，必要な場合に併用できるようにすればよい。

以上検索機械に即してのシソーラスの使用上の注意事項を略述したが，実際の使用法に関しては，JICST の



メンバーによる研究，外務省の発表，その他企業体，団体などの具体例の発表を参照されたい。

今回は取り上げることができなかったが，機会を見て，機械によるシソーラスの自動的作成の問題を論じる予定である。この点については，C. T. Abraham が A.D.I. の会議で発表した論文[22] が参考になることを付記しておく。（図書館学科）

1) Sharp, John R. *Some fundamentals of information retrieval.* London, Andre Deutsch, 1965. p. 131.
2) Vickery, B. C. *Classification and indexing in science.* 2d ed. London, Butterworth, 1959. p. 195–203.
3) Foskett, D. J. *Classification and indexing in the social sciences.* London, Butterworths, 1963. p. 177–80.
4) Vickery, B. C. "Thesaurus—A new word in documentation," *Journal of documentation*, vol. 16, no. 4, Dec. 1960, p. 182.
5) Nicolaus, John J. *The automated approach to technical information retrieval:* Library applications. Washington, D.C., Dept. of the Navy, Bureau of Ships, 1964. (NAVSHIPS 250-210-2) 44 p. 参照。
6) Adams, S. and Jaine, S. "Searching the medical literature," *Journal of the American Medical Associaton*, vol. 188, no. 3, Apr. 1964. p. 251–4に加筆し，訂正を加えた。
7) 索引形式その他詳細な点については，緒方良彦氏が，**情報管理** (vol. 7, no. 5) その他に発表されているので参照されたい。
8) Sharp, *op. cit.*, p. 134.
9) Bar-Hillel, Yeoshua. Theoretical aspects of the mechanization of literature searching. 〈*Language and information.* London, Addison-Wesley Pub. Co., 1964〉 p. 355.
10) Bernier, Charles D. and Heumann, Karl F. "Correlative indexes. III. Semantic relations among semantemes—the technical thesaurus," *American documentation*, vol. 8, no. 3, July 1957, p. 211–20.
11) Farradane, J. Fundamental falacies and new needs in classification. 〈*Essays in librarianship in memory of William Charles Berwick Sayers.* London, L.A., [1961]〉 p. 125–35.
12) Black, J. D. "The keyword; its use in abstracting, indexing and retrieving information," *Aslib proceedings*, vol. 14, no. 10, Oct. 1962, p. 353.
13) *Ibid.*, p. 318.
14) *Ibid.*, p. 319.
15) Papier, L. S. and Cortelyou, E. H. "Use of a technical word association test in the preparation of a thesaurus," *Journal of documentation*, vol. 18, no. 4, Dec. 1962, p. 183–7.
16) Jolley, J. L. "The mechanics of co-ordinate indexing and its relation to other indexing methods," *Aslib proceedings*, vol. 15, no. 6, June 1963, p. 161–9.
17) *Ibid.*, p. 162.
18) *Ibid.*, p. 163–4.
19) *Ibid.*, p. 166.
20) Swanson, Don R. The formulation of retrieval problem. 〈Galvin, P. L., ed. *Natural language and the computer.* New York, McGraw-Hill, 1963〉 p. 259.
21) *Ibid.*, p. 261.
22) Abraham, C. T. Techniques for thesaurus organization and evaluation. 〈*Proceedings of American Documentation Institute.* London, Cleaveland-Hume Press, 1964〉 p. 485–97.