

リモート・センシングに関するシソーラスの作成

Compilation of a Remote Sensing Thesaurus

武 田 裕 子

Yuko Takeda

Résumé

A thesaurus in the fields of "remote sensing" was constructed by following the steps shown in the *Documentation-Guidelines for the Establishment and Development of Monolingual Thesauri*, published by ISO in 1974.

There have not yet been published so many research papers on this subject of "remote sensing." However, the subject fields in which this technology is applied are expanding to various areas, such as meteorology, land use, environment, disaster, agriculture, forestry, and hydrology. And there is no published thesaurus that covers all of these fields properly, while the need for building up a data base on the subject is increasing. That was why the compilation of a thesaurus on this subject was attempted.

For selecting the terms to be used as descriptors, *Proceedings of the International Symposium on Remote Sensing of Environment*, vols. 6~10, 1969~1973, were chosen as the sources. The *Proceedings* contained 680 papers on the subject. To make a frequency table of appearance of words in these papers, a KWIC index was compiled.

The *Remote Sensing Thesaurus* thus constructed is not fully completed and has yet many problems to be solved. For instance, the distribution of descriptors to the 13 categories selected for this thesaurus is not well balanced and some certain categories, such as "Technology," "Areas," received quite a number of descriptors, while others, such as "Transportation," "Roads," received none.

The selection of category terms and distribution of descriptors may have to be reexamined.

武田裕子：石川島播磨重工業（株）技術管理部情報担当

Yuko Takeda, Technical Development Group, Administrative Office, Technical Administration Department
Ishikawazima Harima Heavy Industries, Chiyoda-ku, Tokyo.

はじめに

- I. シソーラス作成のガイドライン
- II. リモート・センシング
- III. リモート・センシング・シソーラス
 - A. 作成目的
 - B. 作成方法
 - C. 構造
- IV. 問題点
 - A. 評価方法
 - B. 作成上の問題
 - C. 内容の問題

おわりに

はじめに

シソーラスという語が初めてドキュメンテーションの分野に登場したのは1957年のことであると B.C.Vickery は彼の論文, "Thesaurus—a new word in documentation"¹⁾で述べている。それ以来, シソーラスということばがドキュメンテーションの分野において数多く使われるようになってきた。そして, "現在のようなシソーラスの形式が整ったのは1959年である。"²⁾しかし, それがしっかりと根をおろしたのはさらに2, 3年後で, このごく初期の頃はシソーラスは "概念別に編成された単語集, ABC順であることは少ない"³⁾といったような漠然とした定義しか与えられていなかった。それが更に数年を経ることによって, 驚く程詳しい定義を与えられることになる。T. L. Gillum は1964年の論文でシソーラスについて次のように述べている。

シソーラスは, 索引語をコントロールし, 表示するための一つの道具である。そこでは, 言葉をあらかじめ指定された方式でコントロールし, 語を注意深く選び出し, 明確に定義し, 事物と関係づける。そして, それらのことばを利用者が迅速正確にとり出せるように配列する。⁴⁾

この定義は, 70年代の今日において述べられているシソーラスの定義とほとんど変ることがない。その後発表された定義と共にこれをまとめてみれば, シソーラスとは, ①文献の索引, 検索に用いるもので, ②概念を表すコントロールされたキーワードと, 文献中で使用される

語を結びつけ, ③語と語の概念間の関係を同義性, 階級性, 関連性などの点からまとめたもの, つまり, キーワード辞書である, ということができるだろう。

また, ISO (the International Organization for Standardization) によって出された *Documentation—Guidelines for the establishment and development of monolingual thesauri* ではシソーラスを,

機能に関して——

シソーラスとはドキュメント, 索引者, 利用者の用いる自然語から, より統制のとれた "システム言語" へ移すときに用いられる, 術語統制のくふうである。

構造に関して——

シソーラスは知識の特定領域をカバーする, 語義的包括的な関連用語の, 統制された機能的な語彙集である。⁵⁾

と定義している。

1950年代の後半に "シソーラス" という語が, ドキュメンテーションの分野に登場して以来, 現在までに, 様々な領域において数多くのシソーラスが作成されてきた。そういったシソーラスの中でも特に代表的で, リモート・センシング・シソーラスの作成に当って模範となりそうなものとしては, アメリカの工学者合同委員会 (Engineers Joint Council) と国防省との共同で作成され, 1967年の12月に刊行されたE J Cシソーラスがある。このシソーラスの対象分野は工学とその関連科学で

あるが、特徴的なこととしては、軍事関係の用語が目立って多いことがあげられる。

E J C シソーラスは大きく四つに区分される。

- (1) Thesaurus of terms (ことばのシソーラス)
- (2) Permuted index (順列索引)
- (3) Subject category index (カテゴリー索引)
- (4) Hierarchical index (階層索引)

(1) はこのシソーラスの主体で、ディスクリプタが一連のアルファベット順に並べられており、これにはカテゴリーナンバが付与されている。そしてその下段には、USE, UF, BT, RT, NTの各参照語が与えられている。(2)はシソーラス中のすべての単語のアルファベット順リストである。

次に、INIS (International Nuclear Information System) シソーラスがあげられる。INIS シソーラスは磁気テープにおさめられた、(1)基本シソーラス、(2)基本逆引きシソーラス、(3)展開シソーラス、(4)展開逆引きシソーラス、(5)印刷シソーラス、(6)エラーリスト、(7)自動修正辞書、の七つの辞書から構成されている。印刷シソーラスはアルファベット順に並べられており、各ディスクリプタの末尾に過去約百万件における使用頻度が記入されている。

相互参照としてはBT, NT, RTがあり、その他に禁止語を指示するものとしてUSE, SEE, USE=がある。

MeSH (Medical Subject Headings) は米国国立医学図書館で開発されたMEDLARSの索引、検索、目録作成に使用するためのものであり、理論に基づいて体系的に作成されたものではなく、米国国立医学図書館における各業務、諸活動の伝統と経験から自然的に発生したものであるという、大きな特色を持っている。そして、その対象分野は医学、生物一般である。

MeSHは件名標目及び相互参照をABC順に並べたり

ストと、件名標目を関連あるグループに分けたカテゴリー・リストから成る。ABC順リストには相互参照としてSee (をみよ)、See under (より一般的な～をみよ) See related (関連ある～をみよ) の3種が使われている。カテゴリー・リストには14のカテゴリーがあり(1977年版)、2段階に分けて階層を示すようにされている。

NASA シソーラスは航空宇宙局(National Aeronautics and Space Administration)によって発行された、航空宇宙工学分野を対象としたシソーラスである。このシソーラスは、NASA 科学技術情報システム(NASA Scientific and Technical Information System)における文献の索引及び検索に使用することを目的として作成されたものであり、主体となるシソーラスと、その使用をより有効にするための4種の付録から成っている。シソーラスの見出し語は34のカテゴリーと217のサブカテゴリー(1967年版)に分けられており、相互参照としてBT, NT, RT, USE, UFの記号を使用している。四つの付録とは、

Appendix A……Hierarchical display (階層表示)

Appendix B……Category term listing (カテゴリー・リスト)

Appendix C……Permuted index (順列索引)

Appendix D……Postable terms (アルファベット順リスト)

である。

次に、わが国における大規模なシソーラスとして、JICST 科学技術用語シソーラスがある。このシソーラスは1975年に発行され、対象分野は科学技術一般という広範囲のものであり、この分野においてはわが国初の本格的シソーラスである。

JICST シソーラスの構造自体はE J Cのシソーラスに類似している。相互参照の表示にはUSE, UF, NT, BT, RTを使用しており、スコープノートも使わ

第1表 各種シソーラスの比較

シソーラス名	ディスクリプタ	NT	BT	RT	UF	N T の 例
EJC シソーラス	Computers	19	1	23	8	Analog computers
INIS シソーラス	COMPUTERS	34	0	11	1	analog computers
MeSH	COMPUTERS	3	1	0	0	COMPUTERS, ANALOG
NASA シソーラス	COMPUTERS	50	1	24	1	ANALOG COMPUTERS
JICST シソーラス	計算機装置	22	1	0	0	オフライン装置

リモート・センシングに関するシソーラスの作成

第2表 各種シソーラスのまとめ

名 称	EJC シソーラス	INIS シソーラス	MeSH	NASA シソーラス	JICST 科学技術用語シソーラス
作成機関	Engineers Joint Council アメリカ国防省	International Atomic Energy Agency	米国国立医学図書館	航空宇宙局	JICST
作成目的	情報の流れに沿うあらゆる段階で使用	INIS レファレンス・シリーズの一体系	MEDLARS の索引・検索・目録作成に使用	NASA 科学技術情報システムの文獻に使用	JICST理工学文獻検索ファイルに使用
主題分野	工学とその関連科学	原子力	医学・生物	航空宇宙工学	科学技術
言語	英語	英語	英語	英語	日本語
相互参照の記号	USE	USE SEE USE =	SEE UNDER	USE	USE
	UF		X XU	UF	UF
	BT	BT		BT	BT
	NT	NT		NT	NT
	RT	RT	SEE RELATED XR	RT	RT
語数	23,364(1967年)	約 16,000		12,887(1967年)	33,998(1975年)
初版発行年	1967年	1970年	1960年		1975年
備考		7種の辞書のうち印刷形式は1種のみで、あとは磁気テープに入っている。			

れている。NTでは、概念が狭くなるにしたがって文字をずらして印刷されている。

そして別冊として主題カテゴリー別索引があり、大きく分けて14の分野と約160のカテゴリーに分けられている。

これら5種類のシソーラスにおいて“Computers”という単語がどのように取り扱われているかを、その語の下に示されているNT, BT, RTなどの数に重点をあてて比べたのが第1表であり、作成機関、作成目的、主題分野、相互参照の形などを比べたのが第2表である。

I. シソーラス作成のガイドライン

シソーラスの作成に関しては、国際的指針が存在する。これは前に1度述べた *Documentation - Guidelines for the establishment and development of monolingual thesauri* (ドキュメンテーション——単・国語シソーラスの設定及び開発のガイドライン) というものであり、UNESCO と ISO TC/46 の共同作業によって作成さ

れ、ISO 2788—1974 (E) として規格化されている。この中には、シソーラスの構造や表現方法や作成方法及びメンテナンスの方法が統一された、利用しやすい形で紹介されている。

このガイドラインが対象としている適用分野は広く科学一般であるが、この“科学”の中には、用語の意味内容が各学派によってまちまちであることが多い“社会科学や人文科学は含まれていないようである。”⁶⁾

さて内容を簡単に述べると、例えばシソーラスの構造に関しては、ディスクリプタはできるだけ1語のものが望ましい。しかし、複合した方が明解になる場合には複合し、自然の語順で表現するようにする。

また、ディスクリプタの数は適度におさえるようにして、概念同志の組み合わせによる表現を活用する。

ディスクリプタは名詞形か動詞の名詞的部分でなくてはならない。そして単数形と複数形でどちらを選ぶかということは国内規準に従い(英語では普通、複数形)、両方で意味が異なってくる場合には両方とも採用する。ま

た、形容詞形などの非名詞形も、ある程度は採用することが許される。

綴りは最も広く使われているものを選ぶ。綴り方が複数存在する場合には、参照の形でその両方をシソーラス中に入れるようにする。

同音異義語は限定句を使用して区別しなくてはならない。これはその語のすぐ後に括弧に入れて示す。

ディスクリプタの相互関係については、まず同義関係であるが、同義の語は一つの同じ概念に割り当て、その中の一つを優先語としてディスクリプタに用いる。この優先語を指示するにはUSE参照を用いる。逆に、優先語から非ディスクリプタを示す逆参照としてUSED FOR参照がある。

次に階層関係であるが、これは概念の上位、下位の関係を表す。階層関係の中にも包括関係と部分—全体関係の2種類があるが、この二つを区別して別々に示してもよいし、一緒にまとめてしまってもよい。また、包括関係のみを階層関係とし、部分—全体関係は関連関係として扱うことも分野によっては可能である。

関連関係は密接な関係はあるけれども、階層的でもなく同義でもないというような関係に用いる。このような関連関係とは、例えば、反義語、同格、発生論関係、二つの概念の同時使用、原因と結果、手段の関係、材料関係、異なった種類の類似、などである。

以上述べてきたような大別して3種の関係を表す記号は、このガイドラインでは、優先関係にUSE、UF、階層関係にBT、NT、関連関係にはRTを用いるようになっている。

シソーラスの主要部分はディスクリプタに関する完全な情報を含んでいなくてはならない。その情報は大きく分けて以下の四つで構成されている。

1. 概念表現
2. 付加的情報
3. 概念関係
4. 他の特殊な関係

このようにして構成された主要部分へのアクセスをより完全にするために、シソーラスには補助的部分が必要である。それは、例えば、アルファベット順の索引や体系的なリスト、図式表示などである。アルファベット順の索引には、ディスクリプタのみでなく非ディスクリプタも含まれていなくてはならない。一方、体系的リストはシソーラス中に示されるディスクリプタのすべてと、概念のすべてが含まれていなくてはならない。図式表示とは

ディスクリプタの相互関係を図解して示すことである。

ディスクリプタの選定法は、そのシソーラスの構成、使用目的、作成計画の背景などによって決定づけられるが、だいたい四つのステップがある。つまり、収集、吟味、評価、決定である。このうちの収集に当たってのソースとしては次のものがあげられる。

潜在的な利用者と主題専門家
国際的、あるいは国内的標準専門辞典
カレントな文献
術語に関する論文、あるいはリスト
既存のシソーラスと分類表
術語集
専門雑誌の索引
抄録サービス
テキストブック、ハンドブック、概要書
経験的に資料を索引すること⁷⁾

さらに評価の段階では、次のようなことを参考にして作業を進める。

- (1)文献中に、あるいは、既存の情報蓄積にあらわれる頻度
- (2)検索時の質問中に予期される範囲
- (3)既に容認されたディスクリプタとの関係
- (4)学問分野において、一般に通用する術語としての妥当性と確実性
- (5)特定の概念を暗示したり意味したりする際の、効果と便しさ⁸⁾

このようにして選択されたディスクリプタを一定の形式に従って記録するわけだが、その形式は意図するシステム特有の方式にそって行う。

そして最後に、シソーラスを確定的に作り上げる前に試用を行って、そのシソーラスの適切さ、および有用性を確認することが大切である。

以上がISOのガイドラインのごく簡単な内容説明である。このような、シソーラス作成のガイドラインの標準化の試みは、例えばシソーラス中で用いる記号(NT, BT, UF, etc.)などの表示方法の統一化や、シソーラスの構造の統一化などの形式的な面では非常に効果的であると考えられる。しかし、何らかの形を持ったものを規格化する場合とは異なって、シソーラス作成のガイ

ドラインの標準化をそのままの形ですべて受け入れることは困難であろう。なぜならば、センサーの作成には形式的なことだけではなく、作成手順などの形を持たない要素も多数入っているからである。こういった内容的な事柄は、作成対象となる分野によってそれぞれ事情が異なってくるため、このガイドラインの指示をすべて適用していくことがむずかしい場合が生じてくることも多いと考えられる。

II. リモート・センシング

リモート・センシング(remote sensing),つまり遠隔探査とは、人工衛星や航空機などによって、遠く離れた物体を非常に長い距離において観測する研究方法である。その観測には電磁波を用い、観測の対象となる物体によって、その特性別に、様々な波長のものを使い分けている。

リモート・センシングは現在、非常に多方面の領域に渡って利用されており、その最たるものとして気象観測がある。気象庁の天気予報が気象衛星 NOAA やひまわりによって送られてくるデータに基づくものであることは周知の通りである。その他の方面としては、地質、地形、土地利用、地下資源、水資源、海洋資源、環境保全、農業、林業などに利用され、資源の有効利用や環境、公害問題にとって重要な資料を提供している。

これらのリモート・センシングのデータは膨大なものであり、人間がいちいち詳しく分析することはできず、通常、電子計算機によって処理されている。

リモート・センシングの観測には、前述のように種々の波長の電磁波が用いられるが、電磁波の波長が変わると、リモート・センシングに使用する探査のためのセン

サ(sensor)も変えなくてはならない。たとえば、リモート・センシングのデータを得る方法には、カラーフィルムや白黒フィルムとフィルタを色々と組み合わせて対象物体を撮影する写真方式が広く利用されている。しかし、写真フィルムによって記録することのできる電磁波は $0.9\mu\text{m}$ 前後でしかなく、それよりも長波長域の電磁波によってのみ得られるデータの記録には、光電変換素子を用いたメカニカル・スキャナ方式やレーダー方式が用いられている。

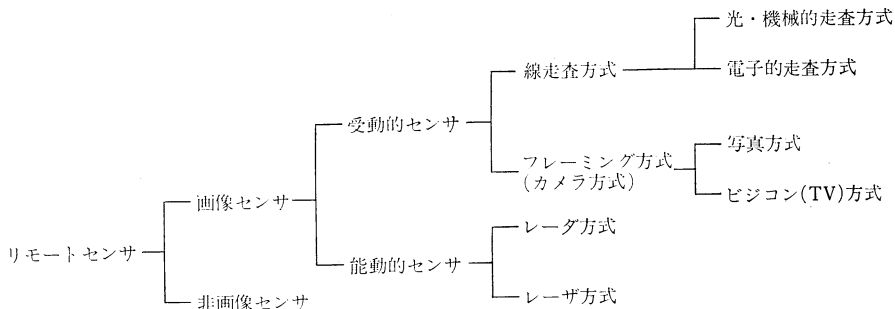
リモート・センサを技術方式によって分類すると第1図のようになる。

画像センサとは字のごとく、観測対象物を画像に結んで分析するものであり、非画像センサは赤外線放射計、マイクロ放射計などを用いるもので、地表物体から放射される赤外線やマイクロ波の放射量を測定することによって観測が行われる。

画像センサのうち、受動的センサは、太陽の光の反射や対象物自身からの自然的な放射を利用するものであり、能動的センサは、対象物に人工的に光を当てて、その反射を利用するものである。また、受動的センサのうちの線走査方式は、主に光・機械的方法によって地上の物体の上を走査する方法であり、近紫外から遠赤外までの非常に幅広いスペクトルの範囲の記録に用いられる。

フレーミング方式(カメラ方式)は、地上の物体から反射するエネルギーを、光学系シャッターを通して、主として蓄積型の感光面上に結像記録するものである。

線走査方式には、光・機械的走査方式と電子的走査方式とがあるが、前者は解像度を高くすると——つまり、森林の大きさを観測することよりも、樹木1本1本を識別するような場合——本体が大きくて重たいものになっ



第1図 リモート・センサの分類

出典：和達清夫〔等〕著 リモートセンシング，東京，朝倉書店，1976年，p. 31.

てしまう。後者は、前者のそのような欠点をカバーするような走査方式であるが、現在、実用化への研究を進めている段階である。

フレーミング方式には、写真方式とビジコン (TV) 方式がある。写真方式は、検出器表面の光化学的变化によって潜像が形成され、それを現像処理するものである。ビジコン方式は、画像がチャージされた分布の形をとり、そこから電氣的に読み出されるものである。

以上がリモート・センサの分類とその簡単な説明であるが、実際の利用面からみると、農業、林業、地質、水質など、リモート・センシングの主要な対象分野における観測は、せいぜい紫外から中間一遠赤外止まりである。そのため、実際に使われるセンサも、 $0.9\mu\text{m}$ 前後においては、一番性能の高い写真方式や光・機械的走査方式が主である。

シソーラスの作成においてディスクリプタを選定する場合、当然その対象分野の広さが問題になるが、リモート・センシングは土地利用、土壤調査、災害地域の調査、樹木の種類や病虫害の状況の調査、水の特性の調査、岩石の分類など様々な分野に利用されている。

III. リモート・センシング・シソーラス

A. 作成目的

シソーラスは、前にも書いた通り、文献の索引や検索に用いるための道具という範囲内で定義されている。そのために、シソーラスの作成対象となった分野は種々多様であるが、大量の文献をかかえている分野であることが多い。これらの分野に比べると、リモート・センシング分野は、その歴史がまだ浅いこともあり、文献量は決して大部のものではない。しかし、それでもなお、リモート・センシング分野にはシソーラスが必要とされている。

普通、文献量がそれ程大量でない分野であるならば、現時点で存在する既成のシソーラスのうちで、その分野をカバーしているものを選んで、それを使用すれば間に合うという考え方が当然出てくる。しかし、リモート・センシング分野の場合、先にも述べたように、研究対象となる範囲が非常に広く、多領域にわたっている。技術的には、航空技術、写真技術、コンピュータによる解析の技術から、農業、林業、水、気象、地理学まで、一見、何の関連も持たないような分野どうしがリモート・センシングという一点によって結びつけられている。シソーラスは、ある一つの分野を対象として作成されているも

のであり、そのため、リモート・センシングがかかわっている領域すべてをカバーしているような既存のシソーラスはみあたらないのである。このような理由から、やはり、リモート・センシングには、その観点から各領域をとらえた独自のシソーラスが必要であると考えられる。

また、別の点からこのシソーラスの必要性を述べるならば、次のような点があげられる。

リモート・センシングには、探査という性質上、観測データがつきものである。この観測データの種類の、リモート・センシングの手法自体が、時と場合によって様々であるため、当然多様なものになってくる。例えば、大きく分けただけでも、磁気テープ、写真フィルム、カードなどの形態に分けられ、さらにその中で細かい区別が生じてくる。

また、こういったデータ類の他に、観測のための資料として、地図の存在も重要である。その種類も観測の用途別に、大きくは世界地図から小範囲では東京都内の地図まで実に様々である。

このように、リモート・センシング分野には、今のところ文献数こそ多くはないが、繁雑さにおいては手に余る程、数数の資料、データが存在する。しかも、これらは時を追って更新されていくものであり、さらには、古い資料やデータも、新しいものとの比較対象として、廃棄し難いものである。そこで、これらのデータ類や地図に関して整理体系が必要になってくる。

また、リモート・センシング・データの特徴とも言えることであるが、こういった様々なタイプの資料は、個々の資料がそれぞれ独立して存在しているというよりも、むしろ、互いに関連しあってこそ有効なものである場合が多い。したがって、これらのデータは単に整理するだけでなく、相互に関係づけをおこない、迅速なデータ利用をめざさなくてはならない。そして、相互に関係づけをするためには、人為的にコントロールされたことばを使用して、そのデータの持つ内容的、外面的特徴を一つ一つ明示していくが必要になる。従って、リモート・センシング・シソーラスを用いて磁気テープや地図に索引語を付与することは、データの利用上たいへん意義のあることだと考えられる。

リモート・センシングは、前にも述べた通り、歴史の浅い学問である。それでもこの研究が始まって以来、すでに10年以上経過しており、研究の方向や領域もそろそろ確定してきている。また、リモート・センシングは利用価値の高いものであり、今後ますます発展していくこと

が予想される。そうなれば当然文献量も急激に増加していくであろう。そのために、リモート・センシング・シソーラスの作成を試みることは、有効であると思われる。

B. 作成方法

リモート・センシング・シソーラスの作成には、前述の通り、ISOのガイドラインを参考にした。しかし、これはあくまでも参考という程度にとどめ、実際には、リモート・センシング分野の持つ特徴と時間的制約を考慮に入れて、なるべく効率の良い方法をとるようにここがけた。

1. 語の収集

シソーラスを作成する手順には、大きく分けて、次のような四つの段階がある。

- 1) 語の収集
- 2) 見出し語の選定
- 3) キーワード関係の決定
- 4) 編成⁹⁾

このうちの語の収集のソースとなる材料については、ISOのガイドラインでは様ざまなものがあげられている。しかし、リモート・センシング分野の入手し得る材料は限られたものになる。例えば、リモート・センシングに関する専門辞典とか術語集は、わかっている限りでは存在しないので、これを使用する方法は除外しなければならなかった。そこで一番手に入れ易く、しかも、ある程度まとまった量のものを使用できるという理由で、カレントな文献を基にした語の収集を行った。

使用した文献は、*Proceedings of the International Symposium on Remote Sensing of Environment* (環境のリモート・センシングに関する国際シンポジウムの会議録)の第6巻から第10巻、1969年から1973年までの5巻分である。語の収集は、この会議録に収録されているリモート・センシング関係の論文の題名からのみ行った。コンピュータを使用して論文題名を構成している各単語をばらばらにして、アルファベット順及び頻度順に並べかえた。しかし、その際、単語が一つ一つ何の脈絡もなく並んでいるだけでは、ディスクリプタを決定する際の評価に非常に不便なので、各単語の前後関係及び関係の深い語や熟語としての存在を知るために、KWIC索引を作成しておいた。なお、最終的に使用した論文の件数は680件であった。

2. 見出し語の選定

次に、このようにして集めた語の中から、リモート・センシング・シソーラスのディスクリプタとして使用するものを選定するわけである。ISOのガイドラインには、ディスクリプタの評価に当って五つの方法が示されている。今回の場合、その中から“検索時の質問中に予期される範囲”の語のみを選別する作業を行うことにした。このようにして、あらかじめ語をふるいにかけおくことによって、他の観点からの選定作業がより能率的になると考えたからである。また、この選別の際に、同時に、選別した語をすべて各分野別のカテゴリーに振り分けておくことにした。このことは、見出し語の選定作業には一見無関係のように思えるが、今後のキーワード関係の決定の作業には是非必要なことであるし、語の意味内容を検討するという見地からも、同時に行うことが十分可能であり、かつ、一番合理的な作業であるからである。その際に使用するカテゴリーは以下のものである。

土地利用 (Land use)
 環境 (Environment)
 災害 (Disaster)
 気象 (Meteorology)
 農林業 (Agriculture and forestry)
 植生 (Vegetation)
 水系 (Hydrology)
 地象 (Geology)
 土壌 (Soil)
 海岸域、海域 (Coastal area and sea area)
 都市 (Urban land)
 技術 (Technique)
 地域区分 (Geographical)

これらのうち“技術”のカテゴリー以外は既成のカテゴリー区分に用いられていたもので、“地表面資源構成例”と“国土利用に関する諸計画の策定に必要と思われる情報項目”の二つの資料を参考にした。これらはリモート・センシングを対象としたものではないが、扱っている分野がほぼ共通であると考えられたので、カテゴリーとしての区分けの情報源としては適切であると考えられた。なお、これらは日本語の資料であり、シソーラスは英語で作成するので選んだ語の英訳が必要となった。“地域区分”に関してのみは、MeSHのカテゴリー(Z)を使用した。

さて、実際の作業手順であるが、まず、KWIC索引を基にして、リモート・センシングに不要と思われる語と

必要と思われる語を選別した。そして必要と思われる語のみ各カテゴリーに振り分けた。この段階で各カテゴリーに振り分けられた語は全部で721語であった。しかし、このような単純な方法だと、各カテゴリーの中に同義語が氾濫することになる。そこで、各カテゴリーに分けられた語の中で、同義とみなされる語が二つ以上存在する場合には、一番出現頻度の高い語を優先語とした。なおその際には、語の収集の時に作成しておいた頻度表を使用した。ただし、同じ単語でも、単数形と複数形では別別の頻度で計算されているので、あらかじめ名詞の一致をしておくことが必要である。

3. キーワード関係の決定

次にキーワード関係の決定であるが、これは、ディスクリプタをすでに各カテゴリー別に分類してあることを利用して以下のような手順で行った。

まず、カテゴリー別に分類したディスクリプタに、コンピュータ処理上に必要なコード番号を付与した。それは次のようなものである。(第2図参照)

A	0 1	0 0 1	0 1
カ アテ ルゴ フリ ア ベ識 ッ別 のた めの	中 分 類 の 識 別	カ 語 数 ゴリ 識別 こと の	同 の 予 コ 備 ドを さ ける ため

第2図 カテゴリー別コード

AERIAL PHOTOGRAPH (P 4021503)
BT REMOTE SENSING DATA
BT TECHNIQUE
UF AIR PHOTO
UF AIRCRAFT PHOTOGRAPH

AIR (B 9900101)
BT METEOROLOGY
RT ATMOSPHERE
IMAGE PROCESSING (P 2000001)
BT TECHNIQUE
NT AERIAL INFRARED

第3図 リモート・センシング・シソーラスの構造

そして「カテゴリー名・NT・ディスクリプタ・コード番号」の順でデータをコンピュータに input して印刷させる。次に、これと逆の形式「ディスクリプタ・BT・カテゴリー名・コード番号」という編成で、もう一度印刷する。このようにして、でき上った印刷形式のリストを参考にして、さらに下位の概念関係を決定していくのである。また、NT, BTの他に、RT, UFやUSE関係も適宜新しく追加して考察を続けていく。

このような作業が一通り完成したら、リモート・センシングの専門家と相談しながら、スコープノートの作成に移ることが必要であろう。しかし、この論文で報告するリモート・センシング・シソーラスは、この点ではまだ未完の状態である。そこで、最後の編成の段階をとりこしてまとめあげたものが、「リモート・センシング・シソーラス案」である。

C. 構造

この「リモート・センシング・シソーラス」は大きく分けて、アルファベティカル・リストとカテゴリー・リ

(カテゴリー)	(カテゴリー・コード)
LAND USE	A
ENVIRONMENT	B
DISASTER	C
METEOROLOGY	D
AGRICULTURE AND FORESTRY	E
VEGETATION	F
HYDROLOGY	G
GEOLOGY	H
SOIL	J
COASTAL AREA AND SEA AREA	N
URBAN LAND	O
TECHNIQUE	P
GEOGRAPHICALS	Z

第4図 カテゴリー・リスト(1)

(J) SOIL	
SOIL MAPPING	(J 0001300)
SOIL MOISTURE	(J 0001400)
SURFACE SOIL	(J 0001800)
SOIL DRYNESS	(J 0300000)
DROUGHT	(J 0302100)
DUST STORM	(J 0302200)

第5図 カテゴリー・リスト(2)

ストの二つに分けられる。

アルファベティカル・リストでは、ディスクリプタをアルファベット順に配列し、その末尾にコード番号を付与している。そして各ディスクリプタごとに、関係語を示している。

各関係語の表示には、ISOガイドラインにならい、USE, UF, NT, BT, RTの記号を用いている。(第3図参照) なお、階層関係は、最終的には4段階にすることを目標にしている。

一方カテゴリー・リストでは、第4図の各カテゴリーのもとに、中分類とその下位語を2段階に分けて、アルファベット順に並べている。(第5図参照) なお、付与してあるコード番号のうち、左側から2桁が00か99のもの(例えば、J0001300やJ990100)は、入るべき中分類のなかった語である。

IV. 問題点

A. 評価方法

シソーラスを作成した場合、それを確定的なものとして実用化する前に、必ず試験的な使用を行って、そのシソーラスを評価する必要があることは前にも述べた通りである。

試験的な使用の方法には、大きく分けて索引作成実験と検索実験がある。索引作成実験は、文献に索引語を付与する段階でのシソーラスの有用性を検査するためのものであり、検索実験は、逆に、文献を検索する際の再現率、適合率をみるためのものである。

索引作成実験の1例として、“保健物理シソーラスの作成とその索引実験”¹⁰⁾では次のようなことを行っている。

実験は10名で行い、JAERI-memoおよびJAERI-report、計10タイトルの抄録について索引語を付与する。索引が完成したら、各人が付与した索引語の中で重複して使用されている語の割合と、多数回索引に使用されたキーワードを調べる。そして最後に、その結果に対する検討を行うのである。

一方、検索実験の例は、“(財)機械振興協会におけるIRシステムの研究”¹¹⁾にみることができる。ここでは、検索質問を論理式に転換して、そのまま検索を行った場合と、シソーラスを使用して検索をした場合との、検索件数の差異及びその中でノイズとされたものの量を調べている。シソーラスの利用によって検索件数が増え、しかもノイズが少なければ、そのシソーラスの有用性が高

いということになる。

さて、“リモート・センシング・シソーラス”であるが、これは一応シソーラスの形にはなったものの、これから徐徐に手を加えていく必要があるものである。そのために、現在の段階で以上に述べたような試用を行うことは困難であるし、無意味であろうと考える。しかしながら、おそらく実用の際に問題になるであろうと考えられる事柄が2, 3生じているので、それらのことについて述べることにする。

B. 作成上の問題

そのまず第1にあげられるのが、語の収集やディスクリプタの選定などの、シソーラス作成の要となる作業を主に機械的な処理によってのみ行ったことである。本来ならばこういった機械的作業に加えて論理的な面からの検討も行われなければ、シソーラスの作成は十分とは言えない。しかしこのような結果を招いた理由は、時間的制約と人手の少なさによるものである。そのために機械的作業が先行することを承知で作成を進めてきたわけであるが、内容的な充実度に不安が残っていることは否定できない。この問題は、今後のメンテナンスや改訂の段階で徐徐に改善していくことができよう。

また、リモート・センシングの専門の辞書がなかったことも問題の一つである。そのために、収集はしたが専門用語であるため、ことばの意味がわからないという語も多く、このような用語は専門家にその意味を尋ねるより他に方法がなかった。そのため、専門用語であるだけにディスクリプタの選定においての重要性は予想されるのだが、概念関係をはっきりとつかみかねて適切な処置のできなかった語も多く、これも今後の善処が必要な問題であると考ええる。

C. 内容の問題

次に内容的な問題であるが、収集したことばを各カテゴリーに分類した時点でディスクリプタに片寄りが生じている。ディスクリプタの振り分けに用いたカテゴリーは、“地表面資源構成例”と“国土の利用に関する諸計画の策定に必要と思われる情報項目”の二つからとったものであることは前にも述べた通りである。しかし、これらの、用意したカテゴリーのうちでディスクリプタが一つも入らないものが数個あった。Ⅲ.C.で示したカテゴリーはすべて、結果的にディスクリプタの入ったものであり、ディスクリプタの入らなかったカテゴリーには次のものがある。

交通、公共施設、道路、地盤

また、各カテゴリーには、その1段階下のものとして中分類が用意されているが、カテゴリー自体にはディスクリプタが入っているが、この中分類のすべてにディスクリプタが振り分けられているわけではなく、空白になっている中分類が目立って多かった。またその逆に、各カテゴリーに分類はしたものの、そのディスクリプタを入れるべき中分類が見当らなかったという場合もかなりある。これらを百分率で示すと次のようになる。

空白のカテゴリー

全体の 23.5%

空白の中分類

中分類全体の 78.9%

該当する中分類のなかったディスクリプタ

ディスクリプタの総数の 33.2%

このディスクリプタの片寄り方を各カテゴリー別に表示すと第6図のようになる。

さて、以上のような問題が生じてきた理由であるが、一つには、語の収集に使用した論文が680件しかなかった

たという、データ量の少なさがあげられる。

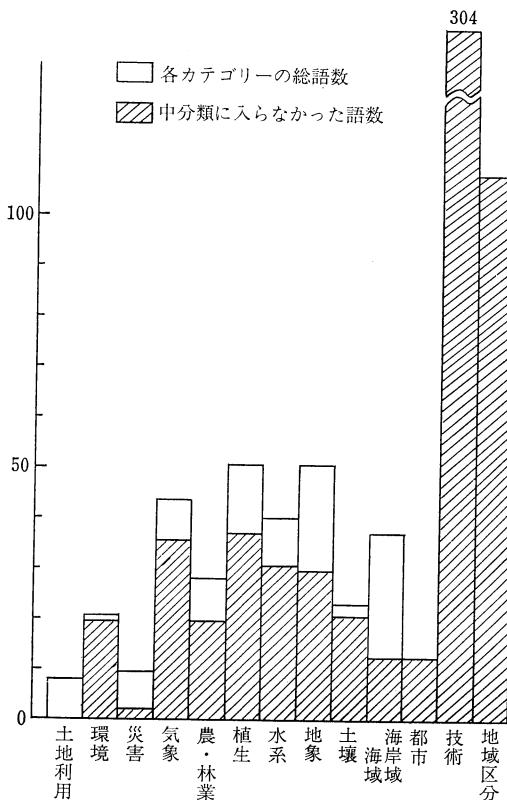
さらに、もう一つの大きな理由として、語の収集に使用した材料そのものの持つ特徴があげられる。語の収集には、リモート・センシングのシンポジウムの会議録を使用した。この内容にはリモート・センシングに関する研究の発表が多い。そしてリモート・センシングの実際の研究は、今回用意したようなカテゴリーのすべてにわたって平均的に行われているわけではなく、どうしても、ある程度片寄ったものになってしまう。その片寄った論文を基にディスクリプタを集めたのであるから、収集されたディスクリプタも、当然片寄りを持ったものになってしまうわけである。

また、研究対象がどのようなカテゴリーに入るものであっても、必ず技術的な事柄がこれに絡んでいるので、他のカテゴリーに比べて、“技術”のカテゴリーに集まるディスクリプタがとびぬけて多くなることになる。

しかし、これらの片寄りに関する問題は、逆に利用の立場から考えてみても同じことが言えるだろう。利用者——すなわち、研究者が自分の研究をしている事柄に関する文献探索を行う場合に、このシソーラスを利用するとすると、探索対象となる研究発表論文にも当然同じ偏向性がみられる筈であるから、ディスクリプタの片寄り、特に技術関係のディスクリプタのとびぬけた多さも、現在の時点では多少やむをえないという気がする。

ここで観点を変えて、別の方向からこの問題を考察してみると、片寄った分類を生じさせる元となったカテゴリーの、選び方自体にも問題があると考えられる。というのは、前にも述べたように、カテゴリーの抽出に使用した資料は、リモート・センシングを対象とした性質のものではない。だから、今回選んだカテゴリーが、リモート・センシングのカテゴリーとして、最適なものとしてしまうのは少し早計であろう。

現在のところでは、集めたディスクリプタの数がそれほど多くないので断定的なことは言えないが、もし今後、ISOのガイドラインに示されているようなさまざまな方法で大量のデータを集めても、なおかつ、ディスクリプタに片寄りがみられるならば、その時点で、余り使用されないカテゴリーを取り除く必要があると考える。そして反対に、これからの研究の進展によって新しいカテゴリーが出現してきた場合には、そのつど、メンテナンスの段階で、新しいカテゴリーを追加していく必要がある。



第6図 カテゴリー別語数グラフ

おわりに

シソーラスは情報検索システムの機械化にとって欠かせない存在であり、現在も様々な分野において新しいシソーラスの作成が進められているはずである。シソーラスの作成は膨大な時間と仕事を必要とし、それはとうてい1人で成し得る事柄ではない。“リモート・センシング・シソーラス”は、その分野について何の知識もなかった人間が、専門家の助言と即席の勉強と即席の手法とで作り上げたものである。内容の不備はもちろんであるが、簡易的に作成されたシソーラスの1例として、また、リモート・センシング分野の文献の処理に関する考察の一端として見ていただければ幸いである。

なお、リモート・センシングの研究はIBMのサイエンティフィック・センターでも進められている。そのために、シソーラスの作成に際してIBMのサイエンティフィック・センターでコンピュータを使わせていただけたことについて、ここでお礼を述べさせていただく。また、リモート・センシングの専門家として、いろいろと指導して下さった、IBMの飯坂譲二氏と、シソーラスの作成に当って惜しみない協力と助言を与えて下さった筑波大学の上田修一氏に深く感謝の意を表する。

(本稿は昭和52年度慶應義塾大学文学部図書館・情報学科卒業論文として提出したものの要点をまとめたものである。)

- 1) Vickery, B. C. "Thesaurus—a new word in documentation," *Journal of documentation*, vol. 16, no. 4, 1960, p. 181.
- 2) 化学索引研究グループ. "thesaurus を用いる索引作成実験," 情報管理, vol. 8, no. 6, 1965, p. 3.
- 3) 坂本徹朗. "検索語辞典," 図書館界, vol. 16, no. 4, 1965, p. 101.
- 4) Gillum, T. L. "Compiling a technical thesaurus," *Journal of chemical documentation*, vol. 4, 1964, p. 29.
- 5) *Documentation-guidelines for the establishment and development of monolingual thesauri*, 1974-03-28, ISO 2788, 1974. p. 1.
- 6) 荒木啓介. "単言語シソーラスの設定と発展のための指針 [その1]—ISO 2788, 1974年版について—," 情報管理, vol. 19, no. 5, 1976, p. 346.
- 7) *Documentation-guidelines for the establishment and development of monolingual thesauri*, *op. cit.*, p. 11.
- 8) *Ibid.*
- 9) 笹森勝之助. "シソーラス," ドクメンテーション研究, vol. 20, no. 1, 1970, p. 17.
- 10) 大井正一. "保健物理シソーラス作成に関する報告," 情報管理, vol. 13, no. 4, 1970, p. 215-26.
- 11) 柴田紘一郎. "(財)機械振興協会におけるIRシステムの研究——シソーラスの作成とそのテストを中心に——[その1]," ドクメンテーション研究, vol. 18, no. 8, 1968, p. 239-47.