

松代群発地震資料のコンピュータによる蓄積と検索

著者	渡辺 一郎, 大倉 博, 尾崎 睿子
雑誌名	国立防災科学技術センター研究速報
号	19
ページ	1-25
発行年	1975-08-30
URL	http://id.nii.ac.jp/1625/00002735/

松代群発地震資料のコンピュータによる 蓄積と検策

渡辺一郎・大倉 博・尾崎睿子
国立防災科学技術センター第3研究部

Information Storage and Retrieval of the Materials of the Research Center of Matsushiro Earthquake Swarm by Using Computer

By

I. Watanabe, H. Ohkura and E. Ozaki

National Research Center for Disaster Prevention, Tokyo

Abstract

Some words prospective to be used as keywords were collected by making reference to various materials and documents which are gathered in the Research Center of Matsushiro Earthquake Swarm.

These words containing Chinese characters cover a wide range of meanings which extends from scientific terms to such words as representing human activities and feelings. By arranging and sorting these words, a thesaurus for information storage and retrieval was prepared.

1. 目的

現在、松代地震センターにおいて、松代群発地震（および関連する他の地震）に関する資料が収集されており、1974年には、その収集数が2,000点を超えた。この数年間、年200点の割合で増加している。この数は、1冊の雑誌も1点としているので、論文数に換算すると6,000点を超えるものと思われる。これらの資料は、表1のように大分類されており、学術的な論文から、小学生の作文、週刊紙の記事にいたる広範囲のものを含んでいる。

さて、これらの資料は、今後の地震、特に群発地震による災害を防止するための参考資料として収集されているのであるから、問題意識を持っている人が、その問題に関する資料をこれらの資料の中から容易に探しだせるようになっていなければならない。実際、松代群発地震が発生したとき、特に群発地震に関する資料を探した所、どこにも整理されていなかったということが、松代地震センター設立の第一の動機なのである。

もちろん、松代地震センターにおいて、資料別あるいは論文別に必要事項を記入したカー

ドを作成し整理しているが、⁽¹⁾6,000枚にもおよぶカードの中から必要なものを探しだすことは容易ではない。表1のような大分類ではなく、もっと細かく分類する方法もあるが、分類を細かくすればするほど、二つ以上の分類項目に関係する資料が増えてしまい、実際には分類ができない状態になってしまうのである。

表1 分類別資料数

年度(昭和)	43	44	45	46	47	48	合計
地震	96	323	76	104	73	101	773
測地地殻変動	53	95	24	14	20	20	226
温泉湧水	33	51	9	3	0	5	104
防災行政	172	142	67	28	21	32	462
その他	177	101	24	32	52	43	429
合計	533	712	200	181	166	202	1994

そこで、コンピュータによる情報検索によって、この問題に対処することが考えられた。

2. 蓄積、検索の考え方

(1) コンピュータを用いる情報検索には、まず、表2に示す二つの仕事が含まれる。

表2のB.3におけるコンピュータプログラムの主要な働きは、各資料につけられたキーワードと、B.1において選ばれたキーワードとが同じであるかどうかを調べることである。

したがって、A.1、B.1において、まったく任意なことばをキーワードとして選ぶと、B.3において適切な資料を探して取り出すことができなくなってしまいます。A.1、B.1においてキーワードとして選ぶことばは、(原則として)定められた範囲内のものでなければならぬのである。このように定められたキーワードの表は、情報検索における一種の辞書とも言えるので、ソーラスと呼ばれている。なお、「原則として」となっているのは、後述するように必要なことばをキーワードとして追加することもありうるからである。

現在のコンピュータにおいて漢字を取り扱うことは非常にむずかしい。したがって、キーワードを表わすのにも英数字を用いる方が良い。しかし、松代地震センターに収集されているものは学術文献だけではないので、英語も用いるわけにはゆかない。また、ローマ字綴りにすると、たとえば振動と震動が同じ表現になってしまうし非常に読みにくい。そこで、われわれはキーワードを漢字を用いて表わすこととした。

(2) A.1の仕事は、資料の内容をわずかなキーワードを用いて表現しようとするのであ

るから、どうしても無理がある。したがって、B.3において必要な資料を探し落さないようにすると、多くの不必要な資料が抽出されてしまい、逆に不必要な資料が少なくなるようにすると、必要な資料を探し落してしまう。必要な資料が取りだせないのは良くないので、不必要な資料が多少抽出されても致し方ないという考え方をするのが普通である。

表2 情報蓄積と検索の仕事(1)

<p>A. 蓄積</p> <p>A.1 資料の内容を適切に表現する若干個のことば(主題語, キーワード)を選ぶ。</p> <p>A.2 各資料のキーワードおよび書誌的事項をコンピュータ内に入れ, 磁気テープや磁気ディスクに記録する。</p> <p>B. 検索</p> <p>B.1 探したい資料の内容を適切に表現するキーワードを選ぶ。</p> <p>B.2 このキーワードとその他の必要な書誌的事項をコンピュータ内に入れる。</p> <p>B.3 A.2において記録しておいたものの中から, B.2の条件に合致するものを探す。</p> <p>B.4 探しだしたものをコンピュータの外に出し, 印刷する。</p>

そこで、B.4において出てきたものを見て、その資料が必要かどうかを人間が判断しなければならぬ。コンピュータによって印字、印刷されたもの(出力)が、人間にとって見やすいものであることも重要なこととなる。このことから、コンピュータからの出力も漢字とすることにした。逆に、このことが、キーワードを漢字で表現することにした理由の一つなのである。

(8) 以上のことから、コンピュータを用いる情報検索において、「シソーラスの作成と改訂」という第3の仕事が非常に重要であることがわかる。

さて、シソーラスに含まれるキーワードの数は有限であり、したがって、キーワードを選ぶためのなんらかの基準が必要である。この基準を定めるのに際して、次の二つを考慮しなければならない。

- (i) どのような資料が集められているか、また今後、集められるか。
- (ii) 探すための質問はどのようなものであるか。

ところが、この両者とも将来の不確定な要素を含んでいる。そこで(i)においては、将来の収集資料を想定するしかないし、(ii)においては、逆にある程度、質問の型を決めておき、質問者にそれを守ってもらい、大きな不都合が生じたとき改訂するという方法をとらざるを得ない。

かくて、「シソーラスの作成と改訂」という仕事は、表3のような流れで行なわれることになる。この報告は、第1報としてC.1からC.3までの経過と結果、問題点について述べたものである。

表3 情報蓄積と検索の仕事(2)

C. シソーラスの作成と改訂
C.1 どのような資料が集められているか調査する。
C.2 検索質問の型を想定する。
C.3 シソーラスを作成する。
C.4 追加される資料、実際の検索質問とその検索結果を調査し、検索質問の型を修正したり、シソーラスを改訂(追加、変更、削除)する。

3. シソーラスの作成

3.1 手順

対象とする資料が地震関係とか原子力関係というような比較的狭い学術文献に限られている場合でさえ、良いシソーラスを作ることはむずかしいのであるが、純粋に学術的な見地からキーワードを決め、シソーラスを作成することができる。

しかし、災害や防災に関する資料の場合には、防災のための行政、災害に際しての人間の心理などに関する資料も対象となり、非常に広範囲にわたるため、キーワードを選ぶことが特にむずかしくなる。そこで、次のような手順により、シソーラスを作成することとした。

- (i) 収集されている資料の一部をよく読み、その内容を適切に表現していると思われることばを選ぶ。このとき、そのことばに対して制限を設けない。適切に表現しているということに重点を置き、ことばを自由に選ぶ。
- (ii) これらのことばにつき、その頻度表を作成したり、いくつかのことばの間の関係を調査したりする。

また、おおまかの分類をする。

- (iii) 上記の結果を参考とし、また検索質問の型を想定しながらキーワードを選び、シソーラスの第一次案を作成する。

資料番号 (調査カード番号)	384						論文番号 (マイクロフィルム整理番号)
地 域	1996	2144	0964				
	篠ノ井						
地 震 名	0844	1142	0398	1427	1178	0931	
	松代群衆地震						
図 名	0844	1142	0398	1427	1178	0931	0705 2214
	松代群衆地震時に						
	0186	0536	2244	2243	2223	2203	1853 0820
	学校のとらえ方						

CH 26

図1(2) 調査用紙(2)

3. 2 資料調査

まず、図1(1)~(4)の調査用紙を作成し、各資料(論文)ごとに必要事項を記入した。図1には、例としてある資料についてその内容が記入されている。図1のような項目としたのは参考文献(1)にしたがったためである。各項目の意味、形態分類や地震分類などの項に記入されている記号の意味についても参考文献(1)を参照されたい。この報告の主題から離れるので説明を省略する。

ここで問題となるのは、「キーワード」の項目の所である。厳密に言えば、今ここに記入されているのはキーワードではなく、キーワードの候補、あるいはキーワード決定のための補助語ともいべきものである。今回の調査において、図1の調査用紙が541組作成された。

3. 3 統計、整理

キーワード候補語についての整理結果の1部を表4に示す。

表4 統計結果

この表4からすぐわかるように、キーワード候補語の数が非常に多く、しかも頻度1というのも多数存在する。これは、内容を適切に表わすということを第一義に考え、制限なしにことばを選んだのであるから当然なのであるが、表5に示すような組み合わせられたことばも含んでいるからである。このような候補語からキーワードを選ぶ考え方については3.4項において述べる。

項 目	数
延 語 数	4 2 4 4
種類の数	2 4 9 8
最大頻度	6 3
頻度20以上	1 0
頻度10以上	4 1
頻度 5以上	1 3 3

さて、表4のような統計をとるのに、すべて人間の手を借りるのはまことに大変である。できるだけコンピュータを使って労力を減らすことを試みた。それには、図1の調査用紙の内容をすべてコンピュータ(の磁気テープ)に記憶させるのがよい。そうすれば、いろいろなコンピュータのプログラムをすることにより、表4のような統計を作成することが容易になる。

図1の調査用紙の内容をコンピュータに入れるには、漢字を取り扱わなければならない。このために次の二つの方法を用いた。

- (i) われわれは、一つの漢字を4個の数字を用いて表現する方法を用いている。たとえば、「避難訓練」は、「1474 1364 0393 1828」となる。そこで、この数字系列をカードや紙テープにせん孔してコンピュータに入れる。もちろん、キーワード候補語であるか、地域名であるかなどの区別がつくようにしておかなければならない。
- (ii) われわれはタブレット型の図形入力装置を持っている。タブレット上の1点をペンで押えると、その点のX、Y座標がコンピュータに送られるものである。このタブレット上に図2のような漢字を記入したシートを載せ、入れるべき漢字をペンで指すことにより、より簡単に漢字情報をコンピュータに入れることができる。

キーワード候補語の間の関係を調べたり、大まかな分類をするのにも、コンピュータを用いることができるが、まったく自動的に行なうというわけにはゆかず、またプログラムを作成するのが大変なので、人手で行なうこととした。すなわち、キーワード候補語一つを一枚のカードに記入し、関連あるものを集め、また分類を行ない、キーワード選定の準備とした。

なお、上記のようにしてコンピュータに入れた資料の情報を、そのままカーブプロッタを用いて表示したものが図3である。検索の結果として印刷されるものも図3の形である。

3. 4 キーワードの選定

以上の準備の後、キーワードの選定、すなわちシソーラスの作成を行なった。そのすべての経過について述べることは、記述が混乱するので避け、ここでは現在の時点での選定の原則、基準および、そのような原則を採用した理由などにつき若干述べることにする。

3. 4. 1 検索する側から見た考え方

検索する側から見るときには、必要な資料を取りだすのに適した質問をすることができるようになっていなければならない。このような質問の様式としては、古くから良く言われている、表6のようなものが望ましい。このほか、「なぜ」、「どのようにして、いかに」、「(活動の)結果はどうか」なども検索の対象となるのであるが、複雑となるので、ここでは表6のものだけに限定することとした。

表6のやり方には、次のような欠点がある。たとえば、「銀行」ということばをキーワードとして選んだとしよう。このとき、「地震の銀行に対する影響」という内容の場合には、「銀行」というキーワードを表6の‘対象’の所に書く必要があり、「銀行において、住民が死傷した」という内容の場合には、‘場所’の所に書かなければならない。すなわち、検索の場合にも、資料へのキーワードづけの場合にも、そのキーワードが主語なのか、対象なのか、場所を示すのか、を区別しなければならないのである。

検索の場合には、その手間はそれほどでもないが、キーワードづけの場合には、この区別をする労力は大きく、非常な負担となる。もちろん、表6のような方式の利点は、この欠点を補って余りあるという考えもあるが、われわれは、キーワードづけの労力を重くみて、表6の方式を採用しないこととした。すなわち、キーワードづけの場合も、検索の場合にも、キーワードを並べるだけとする。表6の例1は、「大震時、長野市、銀行、農民、融資」、例2は、「地震衰退期、皆神山、地震計、観測結果」となる。

3. 4. 2 キーワードの組み合わせ

表6の考え方を採用しないとした場合、表5のようなキーワード候補語をどのように扱うかが、次の問題となる。

(1) たとえば、「傾斜計の機構」ということばをキ

表5 キーワード候補語の例

イ	地下水と地震
ロ	傾斜計の機構
ハ	月別地震回数
ニ	大震時食糧供給計画
ホ	商工業者地震対策特別融資
ヘ	精神的疾患への地震の影響

資料番号	00384	論文番号	
形態分類	C	格納場所	
地震分類	G1	論文別カ-ド	無
時期 始		関連資料	無
時期 終		合併, 廃棄	無
発行時期	1967		
発行者	長野県篠ノ井高等学校		
地域	篠ノ井		
地震名	松代群発地震		
題名	松代群発地震時に学校のとつた措施		
キーワード	退避方法 防災措置 防災訓練 生徒登下校時の心得 砂袋 消火用水 ストープ固定 スラックス着用 災害にそなえる心得 防火計画 高校		

長野県篠ノ井高等学校

図3 印刷結果

ワードとして採用したとしよう。すると、ひずみ計の機構、地震計の機構、重力計の機構、なども当然すべてキーワードとしなければならないであろう。「家屋の崩壊」、「松代群発地震の概況」などについても同じである。このようなものをキーワードとして採用すると、キーワードの数があまりにも多くなりすぎる。

表6 検索質問の型の例

例1		
いつ	大震時に	時期
どこで	長野市で	場所
だれが	銀行が	主語
～を使って		手段
～に対して	農民に	対象
～をした	融資した	活動
例2		
～のときの	地震衰退時の	時期
～における	皆神山における	場所
～の	(地震の)	主語
～による	地震計による	手段
～に対する		対象
～	観測結果	活動

したがって、上記のようなものは、原則としてこれを単一のことばに分解し、それぞれをキーワードとする。ひずみ計、地震計、家屋、松代群発地震、そして機構、崩壊、概況などをキーワードとするわけである。もちろん、これにも例外がある。たとえば、「地磁気変化」を「地磁気」と「変化」に分解するのが原則であるが、「地磁気変化」ということばはよく使われるので、これをキーワードとして採用する。

さて、「機構」とか「概況」などをキーワードとしてよいかどうかは問題のある所である。また、「傾斜計」と「機構」という二つのキーワードではなく、「傾斜計の機構」ということばを、資料の内容を示すものとして採用したい場合もある。これらの問題の対策については後述する。

(2) 「地下水と地震」ということばについても、(1)と同じく、「地下水」、「地震」とい

り二つのキーワードに分けるべきであろう。

コンピュータによる検索においては、「地下水と地震という二つのキーワードを両方とも含む資料を探しだせ。」という検索を行なうことは容易なので、上記のように分解することの問題点は少ない。

ただ厳密にいうと、地下水と地震の関係についての資料か、地下水と地震の両方について書いた資料かを区別することができなくなる。この問題の対策については後述する。

なお、実は、この(2)の記述は、上記のような検索質問を許すことをすでに想定していることに注意すべきである。

(3) 「月別地震回数」ということばに対する方案は、さらに困難である。当然、このことばをそのままキーワードとして採用するわけにはゆかない。日別、時間別、季節別など多くのことばを接頭辞として持つことばをすべてキーワードとしなければならなくなるからである。

「月別」と「地震回数」に分割する案も、「月別」をキーワードとして採用できるかどうか、「機構」などよりさらに問題があるので採用できそうもない。ましてや、「月」というキーワードは、「月別地震回数」ということばからあまりにも離れてしまう。

従来の世の中のソーラースにおいては、この場合、「地震回数」だけをキーワードとして採用している。実は(1)の場合でも、「機構」などはキーワードとして採用されないのが普通なのである。しかし、われわれは、「月別地震回数」などを資料の内容を表現することばとして使いたいという要求が比較的多いということ想定し、次のような計画を立てている。

「月」、「地震回数」ともにキーワードとして採用する。そして、「月別地震回数」をキーワードとして採用しないが、これを表現したいときには、二つのキーワードを接続する特殊記号(たとえば「-」)を使い、「月-地震回数」とする。すなわち、資料の内容を表現するものとして、このような変形をゆるすことにする。

検索の場合、「地震回数」というキーワードを含む資料を検索したいという質問の場合には、「月-」を無視して、「地震回数」、あるいは「月-地震回数」というキーワードを持つ資料をすべて抽出すればよいから問題はない。

一方、「月-地震回数」という検索質問の場合には問題が起る。このとき、「地震回数」とか、「年-地震回数」というキーワードを持つ資料が抽出されることは望ましくない。「月-地震回数」を持つ資料だけが抽出されるのが良い。

しかし、キーワードづけをする人に対して、月別地震回数について書いてある資料には、かならず「月-地震回数」というキーワードをつけ、単に「地震回数」とつけてはならない、と要求するのは無理なことである。したがって、「月-地震回数」という検索質問の場合に、「月-地震回数」を持つ資料だけを抽出すると、月別地震回数について書いてある若干の資料が抽出されないことが起る。

もちろん、このような落ちがあっても致し方ないとか、キーワードづけをする人に上記のような負担をかけるべきだ、という考え方もあるが、われわれは、次のように、検索者に負担をかける案を採用することとした。

「月-地震回数」という検索質問の場合には、'月-'を無視する。したがって、「地震回数」、「月-地震回数」、「年-地震回数」などを持つ資料がすべて抽出される。検索者は、当然、「年-地震回数」を持つものを捨てるし、「月-地震回数」を持つ資料は目的のものである。「地震回数」というキーワードを持つものについては、まずその資料につけられた他のキーワードから必要なものかどうかを判断し、必要と思われるもの、不明のものは、原資料をとりよせて見ることになる。

この案は、キーワードづけする人が、「月-地震回数」というキーワードをつけるように努力すれば、それだけ検索者の負担が減少するという良い面も持っている。

さて、以上のような考え方をさらに拡張して、表7のように種々の異なった記号を用いるならば、キーワードの間のいろいろな関係を表現することができる。ただし、可能であるからといって、ただちにこの考え方を採用してよいということにはならない。資料にキーワードをつけるときの労力が増加するし、検索のための所要時間も大幅に増加するからである。

表7 特殊記号による関係づけ

こ と ば	関 係	キーワード
月別地震回数	～別～	月-地震回数
傾斜計の機構	～の～	傾斜計+機構
地下水と地震	～と～の関係	地下水*地震
地震による湧水	～による～	地震→湧水
農民に対する融資	～に対する～	農民↑融資

したがって、この考え方を採用するかどうか、また採用するとしてもどの程度とするか、現在検討中である。検索プログラムの複雑さがどうなるかも検討事項の一つとなる。実は、この考え方を採用するとしたときの準備として、「月」、「機構」なども特殊なキーワードとして用意しておくこととしたのである。

なお、後述の表9における、「一般説明語」の所に集められているものが、この種のキーワードである。

また、表7を見てもわかるように、この考え方は表6の考え方の変形、省略形とみることもできることにも注意すべきであろう。

表7のような方式を、表5の(≡)以下のことばに適用すると、表8の中央の列のようになり、あまりにも複雑となる。現在の所、これらについては、たとえば、表8の最右列のように処

理する計画である。

表8 長いことばの処置

大震時食糧供給計画	大震時+食糧供給 +計画	大震時+食糧供給, 計画
商工業者地震対策特別融資	商工業者+地震対策 +特別融資	商工業者, 地震対策, 特別融資
精神的疾患への地震の影響	地震→精神的疾患 +影響	精神的疾患, 地震+影響

3. 4. 3 分類とコード

(1) シソーラスが作成されると、資料につけるキーワードは、このシソーラスに含まれるものに、原則として、制限される。したがって、資料の内容を適切に表現することばが見付かったとき、そのことばがシソーラスにあるか、なければそれに近い意味のものはないか、また、そのことばを二つ以上のキーワードを用いて代用できないか、などについて容易に調査できるようになっていなければならない。

あるかないかだけを調べるのであれば、キーワードをアイウエオ順やイロハ順に並べておけばよいが、こうすると関連あるものが離れてしまい不便である。将来、新しいキーワードを追加するかどうかを決める時にも、類似の意味を持つものがすでにキーワードとなっているかどうか重要なポイントとなる。

以上の理由により、キーワードを分類することとした。しかし、「分類」にはまた固有の困難がある。どのような分類を採用しても、一つのことばが二つ以上の分類項目と関連し、どちらの項目に含ませるか判断に迷うということが多く起るからである。何回かの試行錯誤の結果、現在の所われわれが採用している分類の項目を表9に示す。

(2) 表9を見てすぐわかるように、地震名、地域名などもキーワードである、というより、われわれは、これらもキーワードとすることに決めたのである。地震名を地震名の項目の所に書くだけでもよいのであるが、特にその地震名が、その資料のキーポイントであるならば、その地震名をキーワードの所に書いて強調すべきであろう。また、ある個人の活動が話題となっているときには、その個人の名前をキーワードの所に書けるようになっていなければならない。

このようにして、たとえば地域名は、キーワードの所にも地域名の所にも書くことができ、組織名は、発行者、所属、キーワードの3個所に書くことができる。

表 9 分類項目

人名
地震名
地域名(国内)
地域名(国外)
地域名(メッシュ法)
記載誌名
組織名(発行者, 所属)
動物, 植物, 物質名
物理学, 化学, その他の学問
地震
測地, 地球物理
地中, 地表
地理
地球周辺(気象)
災害
対象別, 人間の営み
人間
災害対策
情報, データ, 観測, 調査
一般説明語

(3) キーワードとして採用され、シソーラスに載せられたものを削除したり変更したりすると、すでにこのことばがキーワードとしてつけられている資料がいくつか存在するから、これらも変更しなければならなくなる。この作業は、不可能ではないが、非常に労力のかかるものであるから、一度キーワードとして採用したものは、原則として削除、変更しないこととする。

もちろん、前述したように、防災資料の特殊性から、新しいキーワードを追加することは、原則として認めなければならないであろう。

(4) キーワードの削除、変更を原則として行なわないことにしたので、キーワードを英数

字だけを用いて表現しておくことができるようになった。資料にキーワードをつけるとき、この英数字の表現を使うようにすれば、漢字をコンピュータへ入れるという面倒な作業を、大幅に軽減できるわけである。この英数字の表現を(キーワードの)コードという。検索質問においても、このコードを用いる。

もちろん、コンピュータ内には、表10のようなキーワードとそのコードとの対応表を入れておく。資料につけられたキーワードを印字するときには、この対応表を用いてコードから漢字表現を求め、漢字で印字することができる。

表10 キーワードとコード対応表

記載誌	B000000
⋮	
雑誌	BM00000
技術雑誌	BMT0000
自然	BMTA000
科学	BMTB000
⋮	
一般雑誌	BMG0000
アサヒグラフ	BMGA000
毎日グラフ	BMGB000
婦人公論	BMGC000
⋮	
子供雑誌	BMC0000
小学一年生	BMCA100
小学二年生	BMCA200
⋮	
週刊紙など	BMM0000
週刊朝日	BMMA000
朝日ジャーナル	BMMB000
サンデー毎日	BMMC000
⋮	

コンピュータによる処理を容易にするという観点から言えば、コードのけた数はすべて同

じであり、しかも、できるだけ短かいことが要請される。しかし、けた数については、次の諸点を考慮して、普通のものより長く、原則として7けたとした。

- (i) 地震関係者から、地震名のコードには地震発生年を含ませるよう要請されている。西暦年の後半3けたを用いるとすると、地震名であることを他から区別するけたが1けた、地震名の相違を示すけたが1けた以上必要であるから、全体のけた数は5けた以上でなければならない。
- (ii) 記載誌名、特に学術雑誌においては、発行年ではなく、その号数、巻数が問題になることがある。これをコード内に含めるとすると、少なくとも3けた必要となり、地震名と同じく全けた数は5けた以上となる。
- (iii) キーワードの追加が行なわれるから、けた数には余裕がなければならない。

ただし、表9中の「地震」以下の分類項目については、後述するようにキーワードの数を少なくしぼったので、けた数を3けたとした。けた数が異なるものがあることは、コンピュータによる取り扱いの上で不便であるが、入力の労力を少なくすることに重点を置き、なるべくコードのけた数を小さくすることにしたのである。

(5) コードをキーワードの分類と関連づけることも行なう。たとえば、「BMG0000」というコードは、表10に示すように「婦人公論」を意味するが、「BMG0000」は、すべての一般雑誌を意味する。そこで、「BMG0000」というコードを持つ資料を探したときには、このコードを持つ資料だけが抽出され、検索質問内に「BMG0000」というコードがあれば、第1けたから第3けたまでが「BMG」であるもの、すなわちすべての一般雑誌に関する資料が抽出されるようにする。このような検索プログラムを作ることは容易である。

3. 4. 4 コンピュータ処理、キーワードづけをする側からみた考え方

(1) コンピュータ処理の面から言えば、キーワードの数は、できるだけ少ない方が望ましい。コードのけた数も少なくなって入力の労力も減少し、表10を貯えておく記憶容量も小さくなるなど、利点が多いからである。

(2) キーワードをつける人は、まず資料を読み、その内容を適切に表現していると思われらることを選び、そのことばがシソーラスに載っているかどうかを調べる。もし載っていないければ類似のことばが載っていないか調べる。したがって、シソーラスが非常に大きい、すなわちキーワードの数が非常に多いことは望ましいことではない。

図1の調査用紙に記入する仕事の場合でさえ、最低、1資料あたり5分、ページ数が多かったり複雑であれば20分以上かかっている。この仕事の場合には、キーワードの所に自由なことばを書くことができた。シソーラスを用いるときは、シソーラスの中から適切なキーワードを探さなければならない。1資料あたり最低15分の時間を必要とするであろう。シソーラスが大きければ大きいほど時間がかかることになる。

しかし一方、キーワードの数があまり少ないと、資料の内容を表わすことばがシソーラスの中に入ることが多くなり、結局良いキーワードを見つけるのに苦勞することになる。キーワードの数をどの程度とするかが重要な点となるわけであるが、これについては次のように考えた。

表9における、人名から動物、植物、物質名までは、分類がわかりやすく、ことばとその対象との対応も比較的はっきりしているので、少しぐらいキーワードの数が増えても良いであろう。そこで、3.3項の整理結果に若干の必要なものを追加したものを、ほとんどそのままキーワードとして採用することとした。

しかし、その他の分類項目についてはこのようにするわけにはゆかない。たとえば、「三点観測」ということばを取りあげてみよう。これは地震に関する観測であるから、地震の項に分類されるが、観測の項におくことも考えられる。さらに、人間の自然に対する行為であるから、人間と自然との関係を示す項にも分類できる。

この問題点を解決するため、「三点観測」を、地震の項にも、観測の項にも、人間と自然との関係の項にも置いておき、参照できるようにするのが普通である。もちろん、コードは同じにしておく。したがって、見掛けのキーワード数は2～3倍になってしまう。

3.3項の整理の結果、これらの分類項目に属するキーワード候補語として得られたものは約2,500なので、平均2.5倍として、見掛けのキーワード数は6,000個を超えてしまうのである。

そこで、われわれは、これらの分類項目に属するキーワードの数を1,000以下とすることとした。(実は、この1,000という数を選んだ確固たる基準はないのであるが)

現在の所、キーワードの数は約800である。2,500を800と、約3分の1にしたのであるから、後述の表11を見てもわかるように、選ばれたキーワードは個別的というより包括的なことばになっている。したがって、一つのキーワードが意味する内容をシソーラスに併記しなければならないという欠点があるが、800程度であれば、どのキーワードがどこにあるかを覚えることはさほど大変ではないから、シソーラスにおいて一つのキーワードを2, 3個所に書く必要がなくなる。

以上のような原則、基準のもとに作成されたシソーラス第1次案の一部を表11に示す。〔と〕に囲まれたことばは、キーワードの意味する内容の例示である。なお表10も、シソーラスの一部と考えてよい。この第1次案の全体については、国立防災科学技術センター研究資料として発刊する予定である。

表 11 シソーラスの一部(1)

組織名(発行者, 所属)	A 0 0 0 0 0 0
官庁	A K 0 0 0 0 0
総理府	A K A 0 0 0 0
内部部局	A K A A 0 0 0
:	
科学技術庁	A K A M 0 0 0
内部部局	A K A M A 0 0
:	
附属機関	A K A M B 0 0
:	
国立防災科学技術センター	A K A M B E 0
:	
運輸省	A K I 0 0 0 0
:	
気象庁	A K I X 0 0 0
:	
予報部	A K I X B 0 0
予報課	A K I X B D 0
:	
札幌管区气象台	A K I X L 0 0
:	
釧路地方气象台	A K I X L E A
根室測候所	A K I X L E B
:	

表 1 1 シソーラスの一部(2)

地震	E 0 0
巨大地震	E A A
大地震	E A B
中地震	E A C
小地震	E A D
微小地震	E A E
極微小地震	E A F
気象庁震度階	E A G
⋮	
前震	E B A
主震	E B B
余震	E B C
地震群	E B D
地震活動期	E B E
⋮	
地震波	E C 0
初期微動	E C A
主要動	E C B
P波	E C C
⋮	
地震原因説	E H A
地震予知	E H B
⋮	
(測地・地球物理)	
天文測量	F A A
三角測量	F A B
⋮	
地殻変動	F B A
地球潮汐, アースタイド	F B B
⋮	

表11 シソーラスの一部(3)

(対象別, 人間の営み)	W00
農業	WA0
農業活動〔田植, 施肥〕	WAA
農業施設〔農道, 排水路〕	WAB
農民	WAC
農地, 耕地	WAD
農作物, 農産物〔米, 野菜〕	WAE
農具〔クワ, スキ〕	WAF
農業統計〔収量, 農業人口〕	WAG
農業現象〔根ぐされ〕	WAH
農繁期	WAI
農閑期	WAJ
農業土木	WAK
⋮	
行政	WT0
行政機関	WTA
政府	WTB
官公庁, 官公署	WTC
⋮	
人間	H00
⋮	
心理	HC0
心理調査, 心理テスト	HCA
デマ, ウワサ〔怪文書〕	HCB
⋮	
災害対策	T00
消防消火活動〔化学消防〕	TAA
消防署	TAB
消防人〔火元取締責任者〕	TAC
⋮	

表11 シンソーラスの一部(4)

一般説明語	Z 0 0
活動, 行動, 活躍, 出勤	Z A A
作業, 工事	Z A B
注入, 圧入, 混入	Z A C
⋮	
変化, 移動, 傾斜, 下降	Z C A
破壊, 折損	Z C B
爆発	Z C C
⋮	
状況, 概況, 現況, 様相	Z D A
調子, 様子	Z D B
⋮	
方法, 要領	Z E A
見方, 扱い方	Z E B
⋮	
最高, 最大	Z F A
最低, 最小	Z F B
⋮	
年	Z G A
季節	Z G B
月	Z G C
⋮	
地震時, 大(地)震時	Z I A
⋮	
登下校時	Z J A
⋮	
東部	Z K A
⋮	
校内	Z L A
⋮	

4. あとがき

この情報蓄積と検索の仕事は緒についたばかりであるが、シソーラス作りまでが研究的な問題として一つの重要なくぎりであるので、ここにその経過と結果を報告した。表11は、あくまでも暫定的なものである。検索プログラムその他を考慮して、さらに検討を加えたい。

なお、次のような問題点も重要である。

2項において述べたように、シソーラス作りは情報の蓄積、検索のうちのほんの一部の仕事である。表2のBの検索ができるようになるためには、Aの蓄積の仕事、すなわち各資料にその内容を適切に表わすキーワードをつけ、これをコンピュータに入れる仕事を行なわなければならない。

ところが、キーワードづけの仕事は、非常な労力と時間を費やす仕事なのである。3.4.4項において述べたように、キーワードの数を相当少なくしても、1資料あたり最低15分かかると思われ、1人1時間にわずかに4資料である。資料の内容をよく読むという根をつめた仕事であるから、適当に休憩する必要がある、複雑な資料もあるから、1人1日に20資料を処理するのも困難であると言えよう。しかも、内容をよく理解できないと適切なキーワードをつけることができないから、誰にでもできるという仕事ではない。

このキーワードづけの仕事は、情報の蓄積、検索におけるもっとも重大な問題の一つであることを忘れてはならない。

また、資料につけられたキーワードをコンピュータに入れる労力も無視できない。資料の数が多くなると特に問題となろう。さらに、表10の対応表を記憶させるのは、漢字を取り扱わなければならないこともあって、多くの時間と労力を必要とする。キーワードの数が多いとき、また、キーワードの追加が多いとき、この労力が大きな問題点となる。3.3項において述べた漢字処理の方法は、まだ十分に満足すべきものではない。現在この点についても、さらに研究を行なっている。

終りに、3.2項の資料調査に当たって、種々便宜を与えていただいた、青山氏を始めとする松代地震センタの皆さん、この資料調査をわれわれとともに実施し、しかも、キーワードの選択について貴重なコメントをいただいた、気象庁地震観測所の相原氏、国立防災科学技術センタの第2研究部高橋部長、資料調査室の小見波氏に深く感謝したい。

また、3.3項の統計、整理の仕事について、第3研究部、勝山、福井、諸星の諸氏の手をわずらわしたことを記し、謝意を表わしたい。

参 考 文 献

- 1) 高橋博, 鈴木定夫, 相原奎二(1968): 松代地震センターにおける資料の収集・整理, 第5回情報科学技術研究会発表論文集, 63-69.

(1975年5月15日 原稿受理)