

# 経済統計用語シソーラス作成のための 基礎調査

— 農林省「農家経済調査」の場合 —\*

松井幸子

- I はじめに
- II 農家経済調査について
- III 作業方法および過程
- IV シソーラス作成準備作業
- V おわりに

## I はじめに

文献資料の急激な増加にともなうて、社会科学領域においてはこれまで研究活動の一部と見なされてきた、ドキュメンテーション（情報の収集と蓄積、分類と選別、普及と利用の総称<sup>(1)</sup>）を専門的に分担する機構が生まれてきた。ここで行なわれる一連の作業の中心をなすのが情報検索（information retrieval；ある事象がある人にとって特定の意味をもつとき、その事象の知らせを情報といい、……資料などの情報のコレクションから、必要な情報をさがし出すこと<sup>(2)</sup>）であり、これを組織的に遂行するものを情報検索システムと呼ぶ。

情報とは具体的には文献内容のことであるから、その蓄積にさいして検索システムで

原稿受領 1969年12月27日

\* 本稿は「統計用語（農家経済領域）シソーラス作成の準備」（1968年度提出、慶応義塾大学大学院文学研究科図書館・情報学専攻修士論文）の一部を基にしたものである。その作成途上で、一橋大学経済研究所日本経済統計文献センターの細谷新治助教授、浜田宗雄助手、および演習指導の藤川正信教授から数多くの助言を得たことを記して謝意に替えたい。

なお、詳細な作業過程や調査結果の網羅的提示などは省略したので、上記論文を見られるか、筆者に照会されたい。

(1) 国際ドキュメンテーション連盟 (Fédération International de la Documentation) の定義。日本図書館協会編「ドキュメンテーション用語集 邦訳版」1963年、p. 37参照。

(2) 「ドキュメンテーション基本用語集」（日本科学技術情報センター編「J I C S T利用の手引き 資料編」1969年所載）p. 123-4。

使用した用語と、検索システムを利用してそれをさがす人の使う用語との突き合わせ (matching) によって情報検索ははじまるわけである。このために機械検索 (電子計算機その他を利用する情報検索のことで、わが国では IR という用語をこの狭い意味で使用することが多い。) において使用される用語集が、今日では一般にシソーラス (thesaurus) と呼ばれている。

シソーラスの構成は必ずしも統一されてはいないが、通常、一定の順序 (例. アルファベット順) で配列された見出し語 (keyword; descriptor) のもとに、(1) 同義語、類義語、(2) 上位、下位概念、(3) 関連語 (見出し語に対して、(1)(2) 以外の何らかの関係をもつもの、例. Banking—Accounting) を列記するもので、見出し語が検索システムで使用されていない場合は、その代りに用いられている同義語や類義語に案内し、検索システムで用いられている時は、他の同じくシステムで用いられている語との概念のおよびその他の関連を示すことを目的としている。

シソーラスの作成と利用は、従来、機械検索の盛んなアメリカにおいて強力に進められてきた。<sup>(3)</sup> わが国では、外務省、通産省、機械振興協会、日本貿易振興会、日本科学技術情報センターなどが自家用に作成している。その作成作業は、用語抽出対象の選定、用語の抽出 および用語間の関連づけに大別される。わが国におけるこれまでの試みでは、その検索システムの利用者層を考慮することなく、やや安易に抽出対象を決め、用語の抽出や関連づけにあたっては、抽出対象の特性を考察することなく、画一的方法を適用するきらいがあった。その結果、苦勞して作りあげたものが使いにくく、あまり利用されないと言うものも見受けられるのである。

本稿は、こうした傾向を避けて、抽出対象を限定し、その特性を可能な限り反映した構造のシソーラスを作成するための基礎的データを得ることを目的としている。実際の作業は、農林省「農家経済調査」(指定統計第 36 号) の昭和 40 年度版を使用して、そこより用語を抽出し、統計表の構造および統計表中にあらわれた用語 (以下、統計用語と呼ぶ) の諸特性を調査し、それらを生かした形での用語間の関連を表示した一覧表の作成に終わっている。ただし、これまでのところ、統計用語に関するこうした試みはほ

---

(3) Cf. American Institute of Chemical Engineering ed. "Chemical Engineering Thesaurus: A Wordbook for Use with the Concept Coordination System of Information Storage and Retrieval." New York, 1961, and Engineers Joint Council ed. "Thesaurus of Engineering Terms: A List of Engineering Terms and their Relationships for Use in Vocabulary Control in Indexing and Retrieving Engineering Information." New York, 1964.

とんどなされていないので、今回の結果がシソーラス編成のための有効なデータとなり得ているかどうかは疑問である。今後、同様の調査が数多く実施され、それらの比較検討が進められるようになることを望んでいる。

## Ⅱ 農家経済調査について

農家経済調査（以下、農経調と略記する）が、統計用語抽出対象に選ばれたのは、同調査が古くから日本では重要な調査として続いていると同時に、近年のさまざまな農業政策上の要求に従って、急速に調査事項が増加し、項目間の関係が複雑となり、体系的な使用が難しくなっているため、シソーラスを編成し利用することが適当だと考えられたことが主要な理由であるが、副次的には、筆者が個人の手作業で限られた期間内（約1年間）に処理可能なものとして手頃であろうと、一橋大学の文献センターからお勧めもあったことによる。

使用した農経調は、昭和40年度調査で、12分冊として昭和42年に刊行されている<sup>(5)</sup>。この1ヶ年分12分冊で統計表作表のために使用されている用語は延約3,600語、そのうち重複分を除くと約1,200語であり、これをカード化するのに約100時間かかっている。

本調査が、シソーラスを編集し利用することが適当だと考えられるほど複雑だというのは家計的側面と企業的側面とが未分化な家族労作経営である「農家」を調査単位としているその複雑さの反映として、そこから得られる農家経済全般にわたる総合的統計数値を取りまとめるために使用されている記帳方式が、その実態を表現するさいにさまざまな擬制的勘定科目の設定などを必要としているからである<sup>(6)</sup>。

## Ⅲ 作業方法および過程

### A シソーラス作成方法（一般文献の場合）

- 
- (4) 法律用語に関しては、判例を対象とした戸村和夫氏の報告がある。「法律情報の電子計算機利用と方法について」『びぶろす』第19巻第7号、1968年7月、p.1-20参照。
- (5) 農林省農林経済局統計調査部『農家経済調査報告 昭和40年度版』農林統計協会、1967年。12分冊の明細は下記の通りである。なお、第8-10集は隔年調査のため39年度版（1966年）を使用した。第1集 農家経済総合統計、第2集 農家経済価値統計、第3集 農家物財統計、第4集 農家資金動態統計、第5集 農家生計費統計、第6集 農民栄養統計、第7集 類型別農家経済統計、第8集 経営部門別統計、第9集 農家の租税公課諸負担統計、第10集 農家労働統計、第11集 農産物の商品化に関する統計、第12集 農家の農協利用に関する統計。
- (6) 例えば、年度決算において動態計算と静態計算との結果を一致させるために、「農家可処分所得」と「農家消費支出」の差として定義される「農家経済余剰」と、資産、負債の年増減より算出される「純余剰」との間に存在する調整項目をどう解するかで、農経調における貯蓄の概念規定が4通りに異なってくるのである。溝口敏行「農家貯蓄の項目別分析」『経済研究』第19巻第3号、1968年7月、p.214-7参照。

統計用語を対象とした今回の作業の目的が、シソーラス編集そのものにあるのではなく、その編集にさいしてどのような方法が有効かをさぐる点にあることは、1に記した通りである。ただし、一般の文献を対象とするシソーラス編集方法に批判的検討を加えておくことは、それと対比させて、今回採用する方法を特色づけることができ有用であろうと思われる。

すなわち、日本でされている一般文献を対象としたシソーラスの編集は、抽出対象の主題に関しては非専門家の臨時的作業員（あるいは電子計算機）による用語の抽出と、主題専門家による用語間の関連づけと言った分業がされているのが普通である。用語の抽出にあたっては、日本語文献の場合、分ち書きされていないために用語の単位の決め方が難しいので、はじめはできるだけ長く区切って抽出しておき、あとで専門家が意味や頻度を考慮しながら、単位語 (unit-term) への分割と不要語 (stop word; 検索のさい使用されない語) の除去をした上で、単位語間に関係の付与を行なっている。このように用語の抽出とそのあとの作業を別々の人間が独立に行なっているために、あらかじめ文献中に存在していた用語間の関係は無視された上で、のちにあらためて付与されると言うことになっている。しかも、それぞれの段階で作業者の個人差が働き、用語の抽出もれや単位語設定のさいの不統一、関係付与における恣意などを生じさせているのである。

今回の統計用語に関する試みでは、こうした個人差による恣意性をできるだけ少なくおさえる方法を取りたいと考えた。具体的方法および過程は次節以降に述べる。

## B 統計用語の抽出および分析方法

統計用語シソーラス作成のための基礎調査は、統計表名および統計用語のカードへの転記作業と、その記入結果の分析、およびそれらの所産としての用語間の関係を示した一覧表の作成の3段階に分けられる。ここにはそれらの作業の目的と方法を概括しておく。なお、すべて個人の手作業で、できるだけ個人的判断を要しない方法を採用することは先に記した通りである。

カード化作業では、原則として、表上で1つの欄を占めている用語をそのまま1枚の

(7) 通産省および機械振興協会の場合、比較的詳細な作業過程が報告されている。「I R研究会シソーラス部会中間報告資料」(機械振興協会経済研究所編「I R研究会報告書 インフォメーションリトリバルに関する研究(3)——若干のケース——」1966年所載) p. 89 - 96. および「シソーラス作成とトライアウト」(機械振興協会経済研究所編「経営科学に関する調査 インフォメーション・リトリバルに関する研究(5)」1967年所載) p. 5-72参照。

カードに転記し、その上位欄の用語を書き添える。これは表上に見られる用語と用語間の関係があるがままの形式でカードへ写し取ることを目的とするもので、前記の一般文献を対象とする場合に問題となった事柄（用語の抽出もれ、単位語設定の問題、用語間の関係付与のさいの恣意性など）も一応は回避されている。これらはいずれも個人差の生じる余地のない作業なので、もしも周囲の事情が許せば、完全に機械に任せられるものである。

次の記入結果の分析は、機能的分類による統計用語の考察、統計表における表頭欄（横欄）と表側欄（縦欄）のあり方、用語の自動的処理の可能性の検討である。これらはいずれも統計表および統計用語の特性を明らかにし、それらを生かすことのできるシソーラスの構造を求めることを目的としている。

すなわち、統計用語に機能別分類（調査対象であるか、調査事項であるかによる区分）を適用して、農経調におけるそれぞれの用語の特徴をつかみ、あわせて、それが表頭、表側と言う統計表の構造と何らかの対応関係をもっているかどうかを検討する。もしも対応関係が見い出されれば、それはシソーラスの構造にも組込まれることとなる。

ついで、統計用語の表現形式とそれらの表中における体系表示形式にどの程度一貫性が保たれているかを網羅的に調査する。このことは、詳細はⅢ F 1 に記すが、個人的判断を要しないと言う作業上の基本方針から導出されるシソーラスの電子計算機による自動編集のための、前提条件なのである。その結果、形式的一貫性が十分に保たれているとすれば、表上で各用語に与えられている上下関係が概念上の上下関係とも等しいと言う仮定をおいた上での、およびその範囲での、シソーラスの自動編集が可能となる。また、そこに十分な一貫性が見られないとすれば、その理由を調査することによって、表中における統計用語のあり方を明らかにすることができ、それもシソーラス編集上有用なデータとなるはずである。

最後は用語間の関係を一覧できるようにしたリストの作成であり、この場合も、カード化作業と同様に、統計表上に見られる用語間の関係をそのまま機械的にリスト上に写し取ることを目的としている。統計用語に関する今回の一連の調査によって、この方法が一番望ましいと考えられたためである。

### C 統計用語のカード化作業

用語の採録にはすべて 4"×6" のカードを使用した。カードは3種類作成し、記入方

法別にタイトル・カード、ターム・カード、パターン・カードと名づけた。それぞれの記載内容は次の通りである。(第1図参照)

なお、以下、表上の表頭欄にあらわれる用語を表頭用語、表側欄にあらわれる用語を表側用語と呼ぶことにする。

#### タイトル・カード

統計表名と表番号を転記したカードで、表側用語をコード化して併記した。これは農経調40年度版では、表側用語は一定のパターンで繰り返して出現するものばかりで、その種類数も少なく、グルーピングできたためである。コードは共通コード(例えば地域区分のようにいくつかの集にわたって出現するもの)と固有コード(特定の集だけにあらわれるもの)の2種類である。

#### ターム・カード

表頭で1欄を占めている用語を、原則として、1枚のカードに転記したもので、表番号と欄番号(表頭欄に上から順次番号を与えたもの)、および1段上位の欄の用語の前にGを添えたものが記入されている。(Gはgenericの略)ただし、男女区分のように、単独で表側用語と組み合わせても無意味なものについてはカードを作成せず、1段上位の用語のターム・カード上に「S=男女別」のように記入した。(Sはspecificの略で、以下、S区分と呼ぶ。)また表頭用語が表名の数字による区分のみからなっている時も、カードは作成せず、タイトル・カード上に直接その区分明細を記入した。この場合もその前にSをつけておく。

#### パターン・カード

表頭の下の方の欄で、上位欄の用語の変化に対して、何回も同じ組み合わせであらわれる用語群(以下、パターン・タームと呼ぶ。)を表上の形式のままに転記したカード。このパターン・カード番号を上位欄のターム・カード上にも記入する。

これらのカードの枚数その他は次の通りである。

タイトル・カード…………… 155 枚

共通コード 89 個, 固有コード 35 個

ターム・カード…………… 3,612 枚

見出し語の延語数(総数) 3,612 語

見出し語の異なり語数(種類数) 1,242 語

パターン・カード…………… 19 枚

第1図 カード記入例

(1) タイトル・カード

タイトル・カード番号	… T-87	4(65-122) …	{ 所属ターム・カード番号 例：第4集 65-122番
統計表名	… 形態別流入資金		
調査年度と分冊番号	(40-4) T-3		
統計表番号			
	(p. 44) (p. 48) (p. 52) …		{ 下記コードの出現するページ
	4 <sub>1</sub> ・R <sub>5</sub> F <sub>1</sub> ・4 <sub>1</sub> ・R <sub>5</sub> R <sub>51</sub> ・4 <sub>1</sub> ・L <sub>1</sub> …		{ 表側欄の固有コード (数字のみ) と共通コード (英数字)

(2) ターム・カード

		3-33 …	ターム・カード番号
表頭用語	… きゅうり		
調査年度と分冊番号	(40-3) 2-5-2	G = 野菜類 …	上位欄の用語
		P-3-2 …	パターン・カード番号

統計表番号—グループ番号—欄番号

グループ番号とは、表頭欄に上から順に欄番号を与え、第1位欄から、そのもとに所属する最下位欄までを1グループとし、グループ毎に統計表にあらわれた順に番号を与えたものである。

(3) パターン・カード

		P-3 …	パターン・カード番号
パターン・カード番号 (分冊別)	… P-3-2	G = 農作物(品目別) …	上位欄の用語
調査年度と分冊番号	(40-3) T-2 (33回)		{ 統計表番号 (パターン の生じた回数)
	栽培農家数 作付面積 生産数量		パターン・ターム
	(戸) (a) (kg)		

D 統計用語の機能的分類による考察

1. 統計用語の機能的分類<sup>(8)</sup>

文の構造を理解するために、それを主語と述語に分解してみるように、統計表および統計用語の機能を明らかにするために、客体用語と観点用語と言う区分を設けた。

すなわち、統計表において調査対象を示す用語を客体用語とする。これ自体、体系的分類をもつことが多いが、いかに細分されても、直接は統計数値に結びつかないものである。また、調査対象がどのような立場から観察されているかを示す用語を観点用語とする。この場合、調査対象に固有な概念であるか、それとも他の調査対象の場合にも適用可能な概念であるかによって、固有区分と共通区分に分けることにする。この区分は国際十進分類法 (Universal Decimal Classification) において使われ、補助標数を固有補助標数と共通補助標数にしたのにならったものである。

例えば、統計表名の場合、次のように分解する。(a = 客体用語, b = 固有区分, c = 共通区分とする。)

農業地域別 経営耕地規模別 農業生産物 の 販売 と 家計消費  
 c                      c                      a                      b                      b

今回作成したカード上に記載された各種の用語が、この機能的分類に従うとどのようになるかは第1表の通りである。

第1表 機能的分類による統計用語のカード上の記入位置 (その1)

表 上	記入カードの種類	カード上の記入箇所	タームの機能 <sup>注1)</sup>
表 名	タイトル・カード	見出し語	a b c
表 頭	ターム・カード	見出し語	a b
表 頭	ターム・カード	S区分	c
表 頭	パターン・カード	パターン・ターム	b
表 頭	タイトル・カード	S区分	c
表 側	タイトル・カード	共通コード	c
表 側	タイトル・カード	固有コード	b <sup>注2)</sup>

注1) a = 客体用語, b = 固有区分, c = 共通区分

注2) 一部にcもあり。これは経済統計全体としてみれば、共通区分になり得るものが、農経調では事例が少なかったために、固有コードを与えられたためである。

(8) 浜田宗雄「経済統計における言語」(伊大知良太郎, 水田洋, 藤川正信編『社会科学ドキュメンテーションその情報特性と利用』丸善, 1968年所載) p.322-35参照。本稿とは規定も呼称も異なるが、表中の機能により統計用語を分類するという考えは氏に負っている。



以下、農経調より採録した用語にこの機能的分類を適用して、その構造を分析し、それをもっともよく生かせるシソーラスの枠組を工夫するための手がかりを得たいと思う。

## 2. 農経調における客体用語と固有区分

前項に記した通り、客体用語と観点用語のうちの固有区分とは概念上密接な関係にあるが、共通区分は一応それ自体で独立しているので、客体用語と固有区分について本項で取りあげ、共通区分については次項で扱うことにする。

客体用語と固有区分に関して検討しなければならないのは、統計表から採集された用語のうちの、どのようなものを検索用の見出し語として採用し、それらにどのような枠組を与えるかと言う問題と、カード記載の段階では、1欄内の言葉をそのままの形で1枚のカードに転記することによって避けて通ってしまった、見出し語としての単位語の設定の問題である。今回の採録方法では、1枚のターム・カード上の見出し語の位置に客体用語と固有区分が一緒に記入されているものと、2枚のカードに別々に記入されているものの2種類になっているので、この記入形式の違いと上に記した問題とを関連させながら考えてみる。

客体用語と固有区分とが別々の欄を占めている場合の、客体用語としては、農産物や生活物資など、それ自体として体系的分類をもつものがほとんどであったので、その体系性が表上でどのように表示されているかを調査した。かりにその分類項目を上位、中位、下位の3クラスとする。

例 統計表名：固定資産資材の購入数量と価額

統計用語： { 上位——固定資産資材  
                  中位——大農具類  
                  下位——動力耕耘機

このような体系的分類を背景にもつことが容易に看取される12ケの表において、それぞれのクラスに該当する客体用語が表頭欄のどの位置にあらわれているかは次の通りである。（表頭を上から順に第1位欄、第2位欄とする。）

上位	{	第1位欄	3ケ表
		出現せず	9ケ表
中位	{	第1位欄	6ケ表
		出現せず	6ケ表

下位 { 第1位欄 4ヶ表  
第2位欄 8ヶ表

つまり、下位クラス用語はすべて表頭欄に出現しているが、中位では2分の1、上位では4分の1しか出現せず、その他は省略されている。これは統計表上のスペースの節約のためと思われるが、同時に、統計表名と下位クラス用語さえあれば、表の利用上支障がないことを示している。このことは、検索用の見出し語としても、下位クラス用語だけを採用し、上位と中位用語は省略してもよいのではないかということを示唆する。これは検索システムにおいて用いられる見出し語の数の節約につながる重要な問題であるが、別の観点からも同様のことが導き出されたので、以下に示す。

すなわち、下位クラスの1つの用語を起点として、上位クラス用語へとたどっていくと、中位、上位の順で異なり語数が増加するのである。

例 下位：ぶどう

中位：果樹類、果実類、未成園面積

上位：農作物、農産物、農業生産物、特定農業生産物、食品、生活物資

これは、ある特定の作物が、農作物として作付けられ、生長し、収穫され、農産物として市場に出荷され、あるいは自給食糧として消費され、また次の作付けにも用いられると言うように、農業再生産の過程で、次々とその姿をかえていくことを示している。この再生産過程は固有区分（例、作付面積、生産数量、その他）によって表現できるので、客体用語のうち下位クラス用語と固有区分を組み合わせるような構造の検索方式を設計すれば、上位、中位クラス用語は、検索用の見出し語としては、やはり必要ないとの結論になるわけである。

次に単位語設定の問題に関連させて、表頭欄での客体用語と固有区分のあり方をみると、(1)それぞれ別の欄を占めている場合、(2)同一欄内にある場合、(3)同じ1組の客体用語と固有区分が、ある表では別々の欄に分かれており、別の表では同一欄に併存している場合、の3種類になっている。(2)および(3)は、ある調査対象についての特定の観点からの分析が次第に一般化してきた結果が、日本語のもつ膠着性によって影響を受けているものであるが、検索システムにとっては、(1)では客体用語と固有区分との上下関係の不統一、(3)ではどちらの形式を取るかの不安定さが問題となる。これらは用語整理の段階で方針を定めて統一するしかないが、(3)の併存状態にあるものは、それぞれ分難した方が検索用見出し語の異なり語数は少なくなる。また(2)については、すで

に共存状態にあるわけで、これを再び分離するかどうかは、システム設計における単位語設定に関する基本的問題となる。

### 3. 農経調における共通区分

農経調における共通区分は、第1表に見られる通り、表側欄の共通コードと表頭欄のS区分である。本項の目的は、共通区分に属する用語群の内容を調査し、それらの間の体系性や関連性を見い出すことである。

タイトル・カード上に見い出された共通コードは次の6グループである。

R = 地域区分

L = 経営耕地規模区分

M = 経営形態分類

F = 農家分類

Y = 農家の所得分類

A = 年度区分

これらの12分冊における出現状況は第2表の通りである。

第2表 共通コードの出現状況

分冊番号 コード・ グループ	1	7	2	3	4	11	12	5	6	9	8	10
R	+	+	+	+	+	+	+	+	+	+	+	+
L	+	+	+	+	+	+	+	-	-	-	+	+
Y	+	+	-	-	-	-	-	+	+	+	-	-
F	+	+	+	+	+	+	+	+	+	-	-	-
M	+	+	-	-	-	-	-	-	-	+	+	+
A	+	-	-	-	-	+	-	+	+	-	-	-

注 1) 各分冊の内容については、脚注(5)を参照。

注 2) 傾向をつかみやすくするために、分冊番号は順序を入れ替えた。

注 3) グループ内のコードが出現した分冊は+、出現しなかった分冊は-で示した。

すなわち、Rはすべての集に出現しており、農経調における最も基本的な共通区分である。LとY、FとMはそれぞれいずれかが必ず用いられ、全体的傾向としてはLと

Fが多いが、枠で囲まれた部分では逆になっている<sup>(9)</sup>。そして、Lによって農業経営の規模が示され、Yでは所得概念の導入によって農家経済全体の規模が示されている。またFによって農家を一種と二種に分類し、そのうち一種農家を対象としてMが施されている。従って、LYFMはすべて農家集団をいくつかの部分集団に分割するための指標、つまり農家分類の諸形態としてグルーピングできる。それに対し、RとAは、この農家分類とは別の観点である横断面分析と時系列分析に使用される指標である。

従って、農経調における共通区分は、農家分類、地域区分、年度区分と、表頭のS区分を加えて、次のように整理できる。

#### 農 家 分 類

経営耕地規模区分、経営形態分類、一種二種農家分類、農家の所得分類

#### 地 域 区 分

(以下に示すのは、農経調で用いられていた地域区分の種類である。)

全国農業地域区分、都道府県別、経済地帯区分、東日本・西日本別、農区区分

#### 年 度 区 分

#### そ の 他

月別、年令別、男女別、栄養摂取量区分、金額区分、割合区分、時間区分、めすおす別

### E 統計表における表頭と表側

前節では統計用語を機能的に分類して、主としてシソーラスの枠組について検討したので、ここでは、この客体用語と観点用語が表上でどの位置にあらわれるかによって、表頭用語と表側用語の間に、機械検索上、手がかりとなるような差異が認められるかどうかを検討する。

今回作成したカード上での、客体用語と観点用語の固有区分、共通区分の記入位置は第3表の通りである。

すなわち、固有区分と共通区分については、表頭、表側の両欄に出現しているので、表上の位置からだけでは区別できない。客体用語については表頭だけなので問題はなさ

(9) 第1集(農家経済総合統計)と第7集(類型別農家経済統計)では、LYFMのすべてが出現しているが、第1集は第2集以下の各分冊から主要項目を抜粋し編集したものであり、第7集は農経調の中で最も基礎的な第2集(農家経済価値統計)のうち、主要項目について、さらに多くの共通区分によって組替え集計したものであるためである。

第3表 機能的分類による統計用語のカード上の記入位置（その2）<sup>注)</sup>

機能的分類	表上	記入カードの種類	カード上の記入箇所
a 客体用語	表頭	ターム・カード	見出し語
b 固有区分	表頭	ターム・カード	見出し語
b 固有区分	表頭	パターン・カード	パターン・ターム
b 固有区分	表側	タイトル・カード	固有コード
c 共通区分	表側	タイトル・カード	共通コード
c 共通区分	表頭	タイトル・カード	S区分
c 共通区分	表頭	ターム・カード	S区分

注) 第1表を機能的分類を規準として、組替えたものである。

そうだが、固有区分の一部がやはり表頭に出現しているので、客体用語か固有区分かの判別ができない。つまり、表上の位置からだけで、統計用語を機能的に分類することはできないわけである。

そこで、機能的分類からは離れて、表頭と表側にあらわれる用語の種類に何らかの差異が見られるかどうかを、あらためて検討することにした。

農経調40年度版では、表名の変化に応じて、表頭用語はかわるけれども、表側用語は同じものが何回も繰り返して出現しており、異なり語数では、表側用語の方がずっと少ない。そこで、表側用語を規準として、それが表側だけにあらわれるのか、あるいは表頭にもあらわれるかを調査した。

その結果、共通区分に属する用語では地域区分と年度区分が表側のみで表頭にはあらわれなかったが、農家分類ではほとんどすべてが客体用語として表頭にも出現していた。農家分類の場合、使用されている用語は、もともと各統計表において調査対象（客体用語）であったけれども、それが同時に農家経済分析のための共通の視点（共通区分）にもなっていることを示しているわけである。また固有区分では、35の固有コードのうち、12コードが表頭にも出現していた。つまり、統計用語の機械検索にあたって、表頭用語、表側用語と言う区別から出発するのは、あまり有効ではなさそうである。

ただし、共通区分中、地域区分と年度区分が表側のみであることが、他の統計書にも該当するかどうかに興味をもち、農経調41年度版を調査してみた。ところがそこでは地域区分中、都道府県別のみが表側で、他の地域区分と年度区分が表頭にまわり、それに対応した客体用語が表側にまわっていた。つまり、年度区分が表頭で安定しているほかは、客体用語も表頭、表側の両欄に出現し、40年度版よりもさらに用語の位置づけが

混乱していたわけである。<sup>(10)</sup>

以上のように、表頭と表側への用語の配置については、何ら規則的なものが見い出せなかったが、この点に関し官庁統計の製表担当者に照会したところ、一般的に言って表側には表頭の2倍の項目が入るので、その時々々の項目数に応じて表頭と表側への配置を便宜的に決めている。表頭へ地域区分をもってくる1つの理由は、報告書の該当部分をそれぞれの地域に配布するさいに、リコピーを1枚ですますためだ、と言う類の回答を得た。

このことはかなり一般的現象らしく、統計担当者向けのガイドブックである『統計調査の企画と実施』の中で、当時の総理府統計局製表部長の友安亮一氏は、統計表の様式は便利でわかりやすい構造とすべきであると言いながら、「統計表は形式的要素としての表の輪廓と、実質的な要素としてのその中に記入する数字とから成り立っているが、ここに述べるものは形式的要素としての表の輪廓すなわち統計表の様式の問題である。」<sup>(11)</sup>としている。また、統計実務担当者たちの分担執筆による『日本農林統計読本』<sup>(12)</sup>では、統計データの質的側面の解説は綿密だが、表の構成に関しては全く触れられていない。

このように、製表担当者にとっては、表頭も表側も同じ集計区分として用いられ、その形式的統一を意図するか、あるいは作表上の実務的便宜性に力点があるだけである。しかしながら、検索システム側にとっては、ことにその自動化のためには、表頭と表側によって、それぞれに配置される用語のグループを機械的に判別できるような規則（例えば、表頭は客体用語と固有区分、表側は共通区分）が設けられることが望ましい。つまり、パンチ・カード・システムにおいて、その欄番号によって、あらかじめ大きくグループングされているのと同様のことを、統計表でも工夫してほしいと言うことである。

## F 統計用語の自動的処理の可能性

### 1. 自動的処理について

シソーラス編集を、電子計算機によって自動的に行なうことが技術的に可能かどうかは別として、部分的にであれ、望まれていることは確かである。その理由は、シソーラ

(10) なお、本稿の規定による共通区分は、農経調の過去5ヶ年分（昭和35 - 39年度）を通観したところ、おおよその傾向として、37年度までは表頭、38年度以降は表側に出現していた。それが本文中に記したように41年度にはまた37年度以前と同じ表頭に移ったわけである。

(11) 友安亮一「統計調査の企画と実施」農林統計協会、1952年、p.140.

(12) 久我通武編「日本農林統計読本」改訂増補版、葵出版、1966年、500 p.

ス編集に要する労力が、人手によった場合、膨大なものになることである。例えば外務省での経験（収録語数 10,263 語，編集期間約 2 年 3 か月，実質作業総人日数延 2,200 人日）から、「労力を節約し，作業を標準化するためには，編集手順を極力機械化することが望ましい。」と菊池敏典氏等によって力説されている<sup>(13)</sup>。

もちろん，この自動編集は，はじめてシソーラスを編集する時だけでなく，その後の改訂作業までを可能とする方法でなければならないが，このことは，統計用語シソーラスの場合，一般の文献を対象とする時よりも一層強く望まれる。場所的時間的にさまざまに変化する事象を描写する統計用語は，その本性上，きわめて変動しやすいので，一般文献を検索するさいに使用するシソーラスより，一層ひんばんな改訂作業が必要となるからである。

ここで「自動化」と言っているのは，自動抄録法に関連して，対象文献の意味内容に重点をおくものと，それを構成する記号群の操作，つまりその統計的処理に重点をおくものとに二大別されているもののうち，一応後者に近い。すなわち，表上で各用語に配当されている欄の上下関係がそのまま概念体系上の上下関係であるとの仮定をおいた上で，自動化とは，人為的修正を加えないで，表名，表頭用語，表側用語に識別符号（表番号，欄番号その他）を付加したものをデータとして入力し，計算機内で適当に処理することによって，農経調全体にわたる体系的概念図（map）を出力することと規定している。

この計算機内の処理と言うのは，抽出された用語の頻度計算とか，表上で各用語に与えられている上下関係をもととして，検索用の見出し語の決定と，それらの間の関係づけを行なうことであるが，本作業では種々の制約から，実際にプログラムを組み，計算機を動かし，得られた出力を検討することはできなかった。そこで頻度計算なり関係づけなりを電子計算機で行なうと仮定した場合，その成否の決定要因となることが予想される，語形式および体系表示形式（用語の配当されている欄の上下関係）の一貫性がどの程度認められるかを調査することにした。

語形式の一貫性が問題となるのは，計算機で頻度計算を行なうための同一語の識別にはきわめて厳密な一致が要求されるからであり，体系表示形式に一貫性が必要なのは，語形式の場合と同様に，わずかな差があっても，別の構造として扱われ，農経調全体に

(13) 菊池敏典，笹森勝之助，高橋達郎「情報内容の処理」『情報処理』第 7 巻第 6 号，1966 年 11 月，p. 309.

わたる統一的な概念図を得ることが望めなくなるからである。以下、それぞれの調査について述べる。

2. 統計用語の表現形式について

農経調より採録したターム・カード上の見出し語について、まず計算機による配列を想定して、逐次的にアルファベット順に配列した。ついで、統計表名や表上の上位、下位クラス用語を参照しながら、同一語と見なすべきものを一か所にまとめるための配列調整を行なった。(この過程を完全に機械化することができれば、本項に述べる調査は不要なわけである。)<sup>(14)</sup> この配列調整分を整理したのが第4表である。

第4表 語形不一致の理由別具体例および件数

語形不一致の理由	具体例 (要調整分→統一形)	件 数	
		延	異なり
A 語形の部分的相違			
(1) 省略の有無			
a 文脈上の省略 (表名, 上位, 同位, 下位用語によって, 用語の一部が省略されていることが明らかな時。)	事業収入 →農外事業収入	176	36
b 統計書の解説や統計数字によって, 用語の一部が省略されていることがわかる時。	耕耘機 →動力耕耘機	23	9
c その他の理由による省略	自家農業労働日数 →自家農業の労働日数	250	57
(2) 語形の相違			
d 文脈上の相違	家計支出 →家計費	31	5
e 同義語あるいは類義語	100 カロリー 当り 価格 →100 カロリー 当り 価額	32	15
f 限定詞部分の順序不統一	年度始農業固定資本額 →農業固定資本額 (年度始)	19	3
B 語形の全体的相違			
g 同義語あるいは類義語	家計費に占める飲食費の比率 →エンゲル係数	37	13
総 計 <sup>注)</sup>		568	138

注) 語形不統一分の延語数は、全延語数 (3,206 語) の 16 %  
語形不統一分の異なり語数は、全異なり語数 (1,242 語) の 11 %

(14) 実際に英語を対象として、蓄積から検索までを自動化する試みも報告されているが、ただまだ実験段階のものと思われる。Cf. Salton, G. "The Evaluation of Automatic Retrieval Procedures — Selected Test Results Using SMART System." *American Documentation*, vol.16, no.3, 1965.7, p.209-22. (なお、"SMART System" とは "Salton's Magical Automatic Retriever of Texts System" の略称である。)



統計用語の場合、それぞれ調査項目として定義が与えられているわけで、その表現形式もかなりのところまでは統一されているのではないかと期待していたが、異なり語数でも1割以上要調整分が見い出されている。第4表から明らかになったことは、統計用語の場合、用語の一部に省略がおこる割合が相当高いということである。同表の語形不一致の理由を省略の有無によるもの（a～c）とその他（d～g）に分けると、それぞれ次の通りである。

	異なり語数	延語数
省略の有無	102 (74%)	449 (79%)
その他	36 (26%)	119 (21%)
合計	138	568

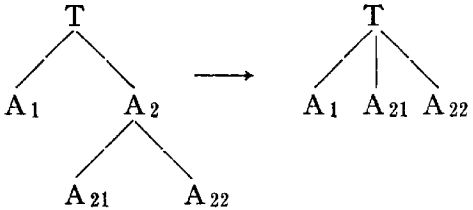
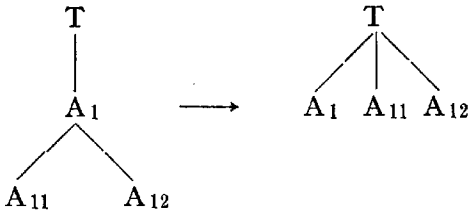
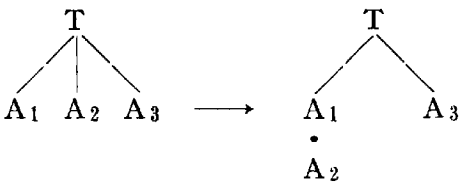
その理由としては、表上ではスペースが限られていると言うことがまずあげられようが、同時に、1つの統計表は1つの論文に相当し、表中にあらわれる用語も、個々ばらばらなものではなく、上下および同位クラスの利用間で互いに補いあって文脈を構成し、各用語の理解を可能にしていると言うこともある。従って用語整理の段階でも、この文脈に注目して、省略部分を再生させればよいと言うことになるだろうが、これを電子計算機に組込むことはそう簡単ではない。かりに調整理由の自動的判別が個別的には可能になったとしても、いくつかの調整理由が1つの用語の中に混在しているものも多いからである。

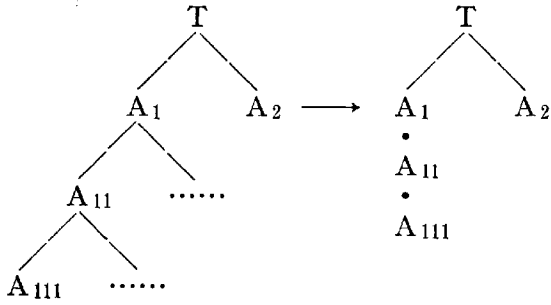
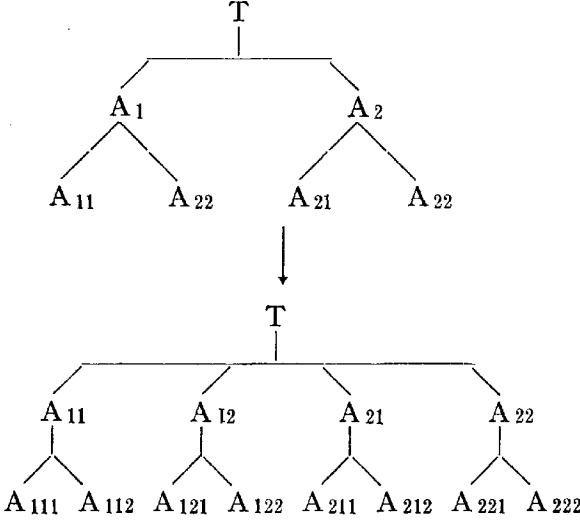
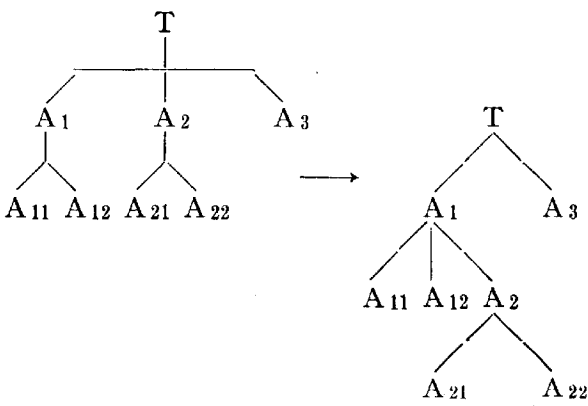
### 3. 統計表の体系表示形式について

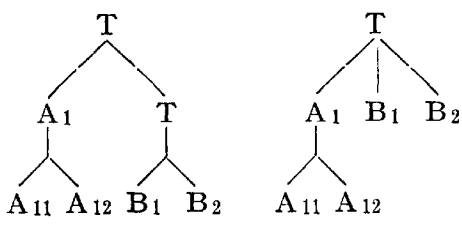
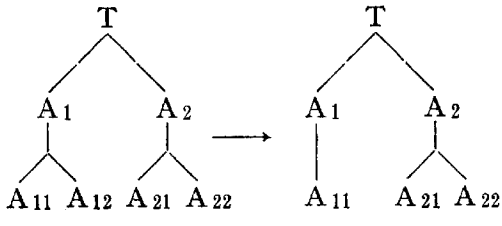
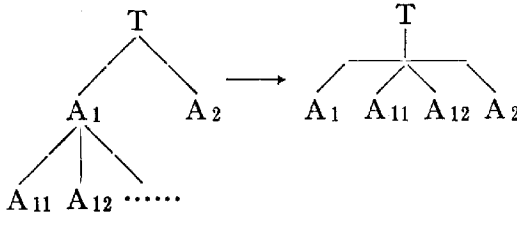
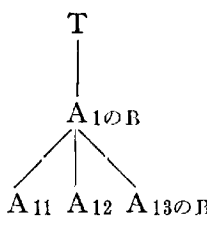
前項における統計用語の表現形式についての調査に続いて、これらの用語によって構成されている統計表上の体系表示にどの程度一貫性が見られるかを調査した。（なお、ここでは、前項で問題とした語形の不統一分は調整されたものとして作業を進めた。）調査は表頭用語（ターム・カード上の見出し語のみ）に限り、表側用語については行なわなかった。農経調40年度版では、表側用語はすべて観点用語からなっており、それぞれ一定のパターンをもっているため、ここで問題とする表示形式の一貫性は保たれているものと見なしたためである。

作業は表頭用語とその表上における欄番号を順次比較し、重複分を除外して、残ったものについて、意味内容の点では等しいが、表上の構成が部分的に異なるものを理由別にグルーピングすることとした。結果は第5表の通りである。

第5表 統計表の体系表示形式不統一分一覧表

不統一理由および件数	具 体 例
<p>(1) 表上で概念上の上位クラス用語が欠けているため、下位クラス用語の欄が繰り上がった時。(25件)</p>  <p>T : 統計表名 A : 表頭用語, 添子の桁数によって, 欄の上下関係を示す。 例. A<sub>1</sub> : 第1位欄, A<sub>11</sub> : 第2欄</p>	<p>1. 経営耕地 2. 畑 3. 普通畑 3. 樹園地 4. 果樹園 4. 桑園 4. その他の園地</p> <p>↓</p> <p>1. 経営耕地 2. 畑 3. 普通畑 3. 果樹園 3. 桑園 3. その他の園地</p> <p>(先頭の数字によって欄番号を示す)</p>
<p>(2) 概念的に上下関係にある用語が, 一部の統計表上で, 同位欄にならんでいる時。(7件)</p> 	<p>1. 農業粗収益 2. 作物収入</p> <p>↓</p> <p>1. 農業粗収益 1. 作物収入</p>
<p>(3) 一部の統計表において, 同位欄の用語が結合して, 語形が変化した時。(18件)</p> 	<p>1. 家族員の労働時間 2. 臨時的賃労働 3. 農業 3. 林業 3. その他</p> <p>↓</p> <p>1. 家族員の労働日数 2. 臨時的賃労働 3. 農林業 3. その他</p>

<p>(4) 他表において上下関係にある用語が、一部の統計表において結合し、語形変化がおこった時。(8件)</p> 	<ul style="list-style-type: none"> <li>1. 常住家族</li> <li>2. 就業者</li> <li>3. 自家農業</li> </ul> <p style="text-align: center;">↓</p> <ul style="list-style-type: none"> <li>1. 家族農業従業者</li> </ul>
<p>(5) ある統計表の用語が、別表でさらに細分されている時。(13件)</p> 	<ul style="list-style-type: none"> <li>1. 農外収入</li> <li>2. 農外事業等の収入</li> <li>3. 農外事業収入</li> <li>4. 林業</li> </ul> <p style="text-align: center;">↓</p> <ul style="list-style-type: none"> <li>1. 林業収入の内訳</li> <li>2. 立木</li> <li>2. 素材</li> <li>2. まき・木炭</li> <li>2. その他の林業収入</li> </ul>
<p>(6) 同じ事柄を扱っているいくつかの統計表において、用語に与えられている欄の上下関係が不統一なもの。(2件)</p> 	<ul style="list-style-type: none"> <li>1. 農業粗収益</li> <li>2. 作物収入</li> <li>3. 稲作</li> <li>4. 水稻</li> <li>4. 陸稻</li> <li>4. 稲わら加工品</li> </ul> <p style="text-align: center;">↓</p> <ul style="list-style-type: none"> <li>1. 稲作</li> <li>2. 陸稻</li> <li>2. 水稻</li> <li>1. 稲わらおよび加工品</li> <li>2. なわ</li> <li>2. わら</li> <li>2. その他の稲わら加工品</li> </ul>

<p>(7) 1つの統計表内において、統計表名が第1位欄で繰り返されたり、繰り返されなかったりして、不統一の時。(1件)</p> 	<p>統計表名： 農業労働力および農業労働投下量</p> <ul style="list-style-type: none"> <li>1. 家族農業従業者</li> <li>2. 専従者</li> <li>2. 補助者</li> <li>1. 農業労働投下量</li> <li>2. 自家農業の労働時間</li> </ul>
<p>(8) 同位欄の2用語が、他表において、そのうちの1用語で代表されている時。(2件)</p> 	<ul style="list-style-type: none"> <li>1. 家計費</li> <li>2. 家計光熱、水道料</li> </ul> <p style="text-align: center;">↓</p> <ul style="list-style-type: none"> <li>1. 家計費</li> <li>2. 家計光熱費</li> </ul>
<p>(9) ある表で上下関係にあった用語のうち、いくつかの下位用語とその上位用語が、別表で同位にならび、もとの上位用語によって、それからもとの下位用語を除いた部分を示している時。(1件)</p> 	<ul style="list-style-type: none"> <li>1. 賃借料および料金額</li> <li>2. 水利費</li> <li>2. 土地改良費</li> </ul> <p style="text-align: center;">↓</p> <ul style="list-style-type: none"> <li>1. 農業経営費</li> <li>2. 賃借料料金</li> <li>2. 土地改良水利費</li> </ul>
<p>(10) 同位欄の用語の形式が不統一の時。(1件)</p> 	<ul style="list-style-type: none"> <li>1. 農業粗収益</li> <li>2. 畜産収入</li> <li>3. 養鶏</li> <li>3. 酪農</li> <li>3. 養豚</li> <li>3. その他の畜産収入</li> </ul>

(11) 不要と思われる項目が、中位欄にある時。すなわち、他に中位用語を含まない上下用語だけの表があり、それらのみで、概念的には十分だと思われる時。(1件)

T

┆

A<sub>1</sub>

┆

A<sub>11</sub>    A<sub>12</sub>

T

┆

A<sub>1</sub>

┆

不要項目

┆

A<sub>11</sub>    A<sub>12</sub>

1. 農家経済の総括

2. 農業所得

2. 農外所得

2. 農家所得

↓

1. 農家経済の概要

2. 所得

3. 農業所得

3. 農外所得

3. 農家所得

ここでも前項と同じく、表上のスペースが限られているために、体系表示の不統一を生みだしている例がほとんどである。これらも理由別に変換して完全形に戻せばよいわけであるが、いくつかの理由が混在している場合、自動化はやはり困難であろう。

#### 4. シソーラスの自動編集の可能性

前2項において、統計表および統計用語について、電子計算機による自動的処理のさいネックとなることが予想される語形式と体系表示形式の不統一分を報告した。この結果、およびⅢ Eに述べた統計表の表頭と表側の問題も含めて、製表側では、統計データの機械検索に対する配慮をほとんど払っていないということが指摘できる。今回の調査は農経調の1年分だけについてであり、さらに調査範囲を広げれば、新たな問題が続出するものと思われる。<sup>(15)</sup>

こうした事態を少しでも改善するためには、機械検索が可能となるために統計表の具備すべき条件について、検索システム側と製表側とが協力して研究を進めることが望まれる。そのための機構を設けるための働きかけは検索システム側から行なわれるべきであらう。<sup>(16)</sup>

ただし、そうした機構が整い、研究が進んだとしても、統計表にあらわれる形のまま

(15) 例えば、指定統計およびその他の主要な統計において、調査が異なると、同一の用語が違った定義を与えられたり、逆に同一の定義が異なった用語で呼ばれたりしていることは、下記の定義集でも通覧できる。行政管理庁統計基準局編「指定統計調査用語定義集」昭37、367p. および、同上編「主要統計調査用語定義集 第2集 定義および分類」昭39、317p. 参照。

(16) ここで検索システムと言っているのは、これは本稿の他の部分にも該当することであるが、一橋大学の日本経済統計文献センターのように、経済統計を収集し、機械検索方式によって一般の利用に供する機関のことである。例えば、官庁関係では総理府統計局が、すでに統計相談所を設けていることでもあり、将来そうした組織の中心となることが期待される。

をデータとする完全自動化システムを開発することは、一般文献を対象とする場合よりは、より可能性があるとしても、やはりかなり困難であろうと思われる。機械検索がかなり進み実用段階に入っているアメリカにおいても、実際に開発され、活用されているのはすべてマン・マシン・システムであって、完全自動化システムは数多く発表されているが、すべてアイデアの域にとどまっているからである。

なお、本節において、完全自動化を念頭において調査を進めたのは、それが統計表および統計用語のあり方を明確に把握するための1つの手段として有効であろうと考えたためである。

#### Ⅳ シソーラス作成準備作業

##### ——KWOC式一覧表の作成——

前章Ⅲで述べた統計用語のカードへの転記とその結果の分析とから、シソーラス作成のための基礎資料として、KWOC (Keyword-Out-Of-Context) 式一覧表を作成することにした。これは、1959年にIBMのH. P. Luhnによって開発されたKWIC索引 (Keyword-In-Context-Index)<sup>(17)</sup>になぞらえた統計用語の一覧表なので、はじめにKWIC索引について簡単に触れておく。

KWIC索引は、一定の長さ（英語の場合、最大60字）の文（普通は論文標題）において用いられている単語のうち、非重要語 (non-significant word) を除いたすべての語をキーワード（見出し語）とし、その数だけ索引を作り、各キーワードが、そのほぼ中央にくるように配置し、キーワードの左右には、それが原文中で占めている位置を示すために、前後の文をそのままならべておく。他の文に対しても同じ操作を施し、最終的には、すべてのキーワードをアルファベット順に配列する。同じキーワードが2つ以上ある時は、それに続く文字を順次配列の対象とする。そして索引の右端にその文献の書誌的事項（著者名、掲載誌名、巻号数、ページ数）をコード化して記載する。

以上がKWIC索引の作成手順の概略であるが、これはあらかじめ非重要語の指定をしておけば、あとは自動的に処理されるわけであり、実際にこの方法で編集された索引誌がいくつもあり、完全に実用の域に入っていると見なされている。<sup>(18)</sup><sup>(19)</sup>

(17) Cf. Luhn, H. P. "Keyword-In-Context Index for Technical Literature (KWIC INDEX)" in Schultz, C. K. ed. "H. P. Luhn: Pioneer of Information Science; Selected Works." New York, Spartan Books, 1968. p. 227-35.

(18) ただし、日本語の場合、あらかじめ分かち書きされていることを前提とする。

(19) Cf. Biological Abstracts Subjects in Context (Biological Abstracts), Chemical Titles (Chemical Abstracts Service).

KWIC索引の長所が、キーワードの意味を、その前後の単語との関連において捕えることができる点、およびその作成が自動化されている点にあることは、開発者Luhnをはじめ多くの人によって説かれて<sup>(20)</sup>いる。そして今回、統計用語の調査のためにKWIC索引にならった一覧表を作成した理由も、これらの2点にあるわけである。

すなわち、ⅢFにおいて、統計用語の形式や統計表の体系表示形式にかなりの不統一が見られることを報告したが、それにもかかわらず、それらが表の利用上ほとんどさまたげになっていないと言うことは、すでに述べたように、統計表においても文脈が存在しているためと考えられるのである。また、自動的に作成できると言う点も、今回のできるだけ個人的判断を要しない方法を採用すると言う基本方針に合致しているわけである。

そこで実際の作業としては、統計表上の最下位欄から順次上へたどり、統計表名までで1つの文脈を構成すると見なして、それぞれの欄の用語をすべてキーワードとし、上述の手順を施すこととした。なお、KWIC索引ではキーワードが索引のほぼ中央に配置されているが、それを左端にもってきて、従来の索引と同じ形態にしたものをKWOC索引(Keyword-Out-Of-Context Index)としているので、今回もこの形式にならうこととした。

このKWOC式一覧表の対象とする用語の種類は、客体用語と観点用語のうちの固有区分であり、共通区分は前にも触れたようにそれだけで概念的に独立しているため除外した。そして客体用語と固有区分のうち、ⅢF3の体系表示形式の検討に先立って重複分を除外した残りのカードをリスト化の対象とした。それはA4版1ページ30行で約270ページ、延語数は数えてないが、異なり語数は1,073語となっている。第2図に<sup>(21)</sup>その記入例を示す。

一覧表上では、見出し語についての1つの統計表の中での位置づけが、その語を中心とする上位欄から下位欄までの用語群によって示され、1つの見出し語のもとの、それらのいくつかの用語群によって、他の統計表の同一語形をもつ語を、その語の上位および下位の用語と共に一覧できるようになっている。

(20) Luhn, H. P. *op. cit.*, p. 228.

(21) 詳細は『農家経済調査KWOC式用語一覧表』として別途刊行したいと思っている。なお一覧表の作成手順は以下の通りである。

KWOC式一覧表の原票として、小片のメモ用紙を用意し、ターム・カード上の見出し語を転記し、欄番号を書き添える。その上位欄と下位欄の用語をそれぞれ同位クラスの欄の用語全部を、上位欄から順次1行ずつに横にならべて書き出す。見出し語の上位用語と下位用語を区別するために、下位用語の行頭は2字分の字下げをする。最下位用語がパターン・タームの時は、パターン・カード番号を記入する。見出し語に対応する表側用語のうちに固有区分がある場合、最下位用語の次行にそのコードを記入する。

第2図 統計用語のKWOC式一覧表記入例

見出し語……………だ い ず	②
(例. 第2位欄)	T 農業生産物の販売と家計消費……統計表名
上位用語……………①	豆 類
(例. 第1位欄)	③ P-3-4, 9 ……………下位用語
	(例. パターン・ターム) <sup>注)</sup>
見出し語……………	②
(同 上)	T 農業生産物の販売と農協利用
	① 農業生産物
	③ P-12-1
	①
	T 農家の消費した生活物資
	② P-5-1
表側 (固有コード) <sup>注)</sup> ……………側	5 <sub>3</sub>
	②
	T 農作物の作付面積と生産数量
	① 豆 類
	③ P-3-2
	①
	T 農産物の商品化
	② P-11-1
	②
	T 食品別の消費量と価額
	① 豆 類
	側 64, 5
	①
	T 特定農業生産物の販売と家計消費
	② P-3-8
だ い ず 加 工 品	①
	T 食品別の消費量と価額
	② 油揚, みそ, 納豆, とうふ
	側 64, 5
だ い ず か す	②
	T 農業生産資材の購入数量と価額
	① 飼料類
	③ P-3-7

注) 記入例中にあらわれた, 表頭パターン・ターム, 表側固有コードの明細は, 以下の通りである。

P-3-2	栽培農家数	作付面積	生産数量
-------	-------	------	------



P-3-4

販 売				家 計 消 費		
販売農家数	数量	価額	平均価額	自給農家数	数量	価額

P-3-7

購入農家数	数量	価額	平均価格
-------	----	----	------

P-3-8

販 売			家計消費	
数量	価額	平均価格	数量	価額

P-3-9

販 売			家 計 消 費		
販売農家数	数量	価額	自給農家数	数量	価額

P-5-1

購 入	自 給	計
-----	-----	---

P-11-1

栽 培 戸 数	栽培農家1戸当平均					販 売 戸 数	販 売 農 家 1 戸 当 平 均 販 売 数 量	全 調 査 農 家 に 対 する 割 合	栽 培 農 家 数 に 対 する 割 合	全 調 査 農 家 に 対 する 割 合
	作 付 面 積	年 生 産 数 内 量	処 分 合 計	販 数 売 量	家 費 計 消 量					

P-12-1

総 額	うち 農協取引額	農協利用率
-----	-------------	-------

5<sub>3</sub>

数 量
価 額

6<sub>5</sub>

6<sub>4</sub>

消費量
価 額

数 量	総 量
	購 入
	自 給
価 額	参考 純摂取量
	総 量
	購 入
	自 給

## V お わ り に

今回の統計用語シソーラス作成のための基礎調査は、前章におけるKWOC式一覧表の作成までで終わっている。ここでは実際に利用可能なシソーラスとするために残さ

れている作業の1つとして、この一覧表を利用して何を明らかにできるかを、用語のカード化にはじまった一連の作業結果と関連させながら検討して稿を閉じたい。

作業方針として一貫して個人的判断を要しない方法を採用したのは、統計表上における各用語とそれらの間に与えられている体系表示形式を個人的解釈で破壊することのないようにとの配慮によるものであり、その結果、KWOC式一覧表においても、各用語を表上においてと同様に見ることができるわけである。従ってこの一覧表を調査することによって、表上で各用語に与えられている欄の上下関係がどのような種類の関係を示しているかを明らかにすることが、当面の作業となる。そこに見い出される関係と、これまでの調査結果とを参考にすることによって、農経調シソーラスの構造もより具体的になってくるはずである。

これまでの調査で明らかになったことは、機能的分類によって統計用語を観察したさいに、客体用語では、農産物や生活物資のように体系的分類をもっているものが多く、それらについては、下位クラスの利用語を検索用の見出し語として採用し、それと農家経済の再生産過程を反映させた固有区分を組み合わせるようにすれば、中位、上位クラスの利用語は検索用の見出し語としては落してしまってもよいこと、および共通区分については、客体用語や固有区分との関係はシソーラス使用時における用語の組み合わせによって表現できるので、共通区分内のグループ別（地域区分、農家分類など）に用語間の体系性や関連性を見い出せばよいことなどであった。

そこで、KWOC式一覧表上で見られる各用語間の関係を分析することによって、客体用語と固有区分との組み合わせで実現される農家経済再生産過程を見出すことにより、シソーラスの構造も決まってくるものと思われる。それは、自動化の検討のさいに触れた農経調全体にわたる統一的概念図ともなるわけである。

そして、このKWOC式一覧表の調査過程において、表上の各用語間の関係について、(1)そのまま自動化可能な部分、(2)一定の修正パターンを用意することで自動化できる部分、および(3)概念体系と一致せず、自動化不可能分、との3つに分けた場合の(3)について、量的データを得ることもできるはずである。(2)についてはⅢFで重点的に調査してあるので、それらを総合することによって、統計用語シソーラスの自動編集についても、ⅢFにおけるよりも、もう少し具体的な判断が下せるであろう。