

SSM職業データにおける自由回答の分析

— SSM職業コーディング支援

エキスパート・システム構築のために —

高橋 和子

はじめに

一般に、社会調査においては、職業は性や年齢とともに基本的かつ重要な属性である(安田・原1982)。とりわけ、社会階層・社会移動に関する実証的研究においては職業が中心的な項目の一つであるため、その代表的な調査である「社会階層と社会移動全国調査」¹⁾(Social Stratification and Social Mobility Survey: 以下SSM調査と略す)では、本人の初職から現職に至るまでの職業経歴²⁾に加えて、配偶者や結婚時における双方の父の職業等も調査項目として設定される³⁾。調査者は正確な職業を把握する必要上、被調査者に直接、職業を回答させたり、調査員の判断で記入させたりせずに、本人の仕事内容や役職、さらに従業先の名称や規模等を一連の職業データとして多元的に収集し、後で調査者自身が総合的に判断して、該当する職業にコーディングするという方法がとられている⁴⁾。

しかし、職業データには選択回答に加えて3種類の自由回答も含まれており⁵⁾、これらを約300個ある職業小分類コードのいずれかにコーディングする作業は、非常に煩雑で膨大なものである(原1993)⁶⁾。

高橋(1994b, 1994c, 1995a)による「SSM職業コーディング支援エキスパート・システム」は、この作業を軽減化する目的で提案された。その概要は、各職業の説明書である『SSM職業分類(改訂版)』(原1993)の内

容をデータベース化し、少数の「職業コーディングの専門家」が職業を決定する際に用いるルールを知識ベースとするシステムと併用することで、効率よくコーディングを行おうとするものである。またコンピュータを利用することでコーディングの一貫性が保証され、信頼性の向上も期待できる。

実際にシステムを構築するには、データベースの設計や辞書の作成等検討すべき課題がいくつか残されているが、これらはいずれも、システムで扱われるデータのうち、主要なものが自由回答であるという特殊な状況⁷⁾に起因するものと考えられる。なぜなら、一般に、自由回答法によりデータを収集すると、被調査者はしばしば質問が要求するものからはずれた回答をすることがあるし、回答に用いられる用語も多様になる傾向を示すからである（高橋1995b）。人手で行えば、必要に応じて他の質問の回答を参照したり、知識を駆使して用語の意味を解釈することが可能であるが、コンピュータはあらかじめ指示を与えておかない限り、そのようなことは行わない。

結局、SSM職業コーディング支援エキスパート・システムにおいては、自由回答をうまく処理することが大きなポイントとなるが、この問題を解決する手がかりとして、今回、職業データが実際にどのような表現で回答されているかについて、1965年SSM調査データを用いて分析を行ったので、本稿で報告する。以下、I章で問題の所在をより具体的な形で明らかにし、II、III章でデータの説明と分析の結果について述べる。最後にまとめを行う。

I. 問題の所在

SSM職業コーディング支援エキスパート・システムにおける主要なデータは仕事内容と役職であるが、前者は自由回答であり、後者も選択回

答ではあるものの「その他」にさまざまな回答が出現する（高橋1994b）。従って、これらは回答をそのままの形でデータベース・システムや知識ベース・システムに用いても、有効ではない。なぜなら、回答は被調査者により日常的なレベルで語られた用語であるのに対し、データベースや知識ベースはある程度、抽象的な用語で表現されたものだと解釈できるからである。これが特に問題となるのは、コンピュータは人間と異なり、同じ意味内容の用語でも、字句が違えば別のものだという判断を下すからである。⁸⁾ 例をあげて説明する。

例えば、高橋（1994b）で用いたサンプル（出典：原1984）に対して人手により職業コーディングを行ったところ⁹⁾、仕事内容が「米作」（サンプル番号003）や「稲」（サンプル番号035）という回答は、いずれも簡単に「076 農耕・養蚕作業者」という職業分類コードに判断できた。¹⁰⁾ しかし、これを本システムで行うとすると、『SSM職業分類（改訂版）』にマッチする用語がないため（穀物という用語はある）、検索ができないことになる。

もし、「穀物」という用語に「米」や「稲」という用語を対応させる辞書を作成しておき、データベースの検索を行う前に通過させれば、マッチングは成功するであろう（図1参照）。このように、人間なら読み替えて解釈するような用語をすべて網羅した辞書を作成できれば、コンピュータもかなりの程度、柔軟な処理ができるはずである。

しかし、実際には、これはメモリの制約や処理時間の関係から不可能である。従って、問題はいかに効率よく有用な辞書を作成するかということになる。もちろん、辞書の充実度がそのままシステムの有効性につながるわけではなく、例えば、回答の用語がマッチした職業分類コードがあっても、次のように単純にそのコードに決定できない場合もなくはない。すなわち、仕事内容が管理になっている場合、「046 管理的公務員」や「051 会社・団体等の管理的職業従事者」等の職業定義文中の用語にマッチするが、役職が課長以上であり、かつ従業先の規模が5人以上でなければ、管

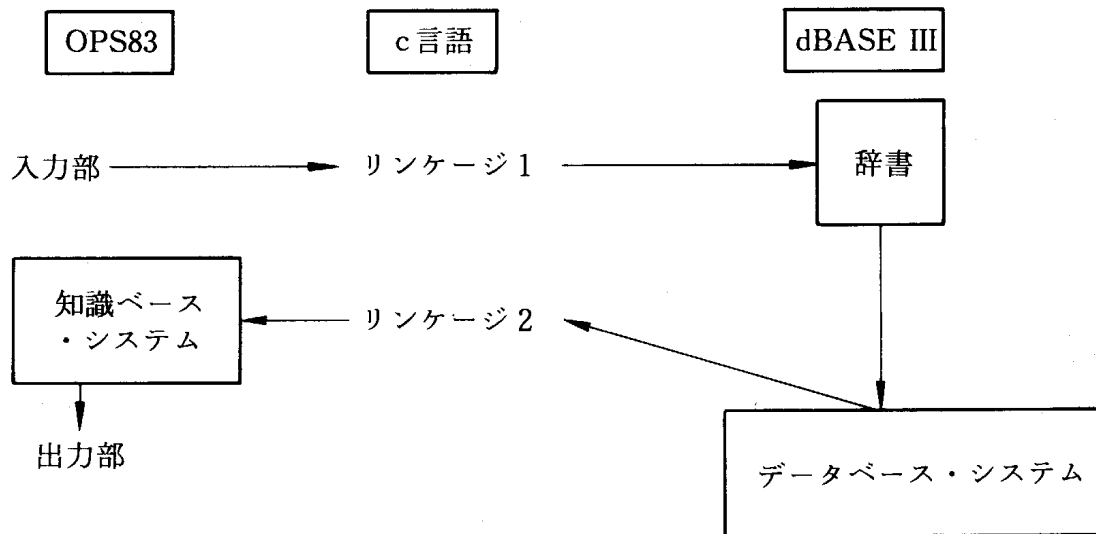


図1 SSM職業コーディング支援エキスパート・システムにおける辞書(高橋1994b)

理的職業にすることはできないというルールが存在する(高橋1994b)。

しかし、本システムでは、戦略的に、まずデータベース・システムにより該当する可能性のある職業を拾い出し、知識ベース・システムにより絞り込んでいく方法をとるために、システムの入り口にある辞書を充実させることが、最低限、必要な条件であると考えられる。

SSM職業データにおける自由回答の処理を行うために、有効かつ効率的な辞書をどのようにして作成すればよいか。この問題を検討するには、トップダウン的には、『SSM職業分類(改訂版)』の内容を調査し、用いられている用語を詳細化、具体化する必要がある、ボトムアップ的には、できるだけ多くの回答を集めて、実際にどのような表現が用いられているのかを知る必要がある。前者については、既に高橋(1994a)で内容の構造化と用語についての検討を行っているため、本稿では後者についての検討を行うこととする。その際、両者を独立に検討しただけでは意味がなく、関

連付けが必要になると思われる。

ところで、本稿における回答の分析は以下に述べる点からも重要である。1つは、本システムの有効性の再確認という点である。前述したサンプルには、明らかにデータベースで検索できないものも含まれており、例えば、サンプル番号011は仕事内容が「いっさい」であり、役職は無回答であった。これは人間でも解釈が不可能であり、従業先の名称や事業内容等を参照しても、おおよその判断を行うことしかできない。このようなものの手順をコンピュータに指示することは非常に困難で、本システムでは判断不能となる。もし実際の回答にこのような表現が多出する傾向があれば、コンピュータによる支援は諦めざるを得ず、ここでその確認を行っておく必要があるだろう。もう1つは、表現内容を把握することはヒューリスティックな知識獲得の増大につながり、知識ベースの構築に役立つという点である。

Ⅱ．データと分析方法

1. データ

分析の対象となったデータは、1965年SSM調査（第2回）における職業データで、調査地点はいずれも東京都23区内である。これは入手の容易さにより選ばれたが、次のように、調査年が古いことによる欠点がある。1つは、データの形式が最新版である1985年調査（第4回）と若干異なる点である。これは、分析の際に適当にデータの項目間移動や削除を行うことで対応できると判断した（2の（3）参照）。もう1つは、産業構造の変化によるデータ自体の古さであるが、これは方法論的には問題はない。ただし、1965年以降に登場する新しい職業に関わる用語の欠落があり得ることを留意する必要があるだろう。

SSM調査の原票には、職業データ以外にも多くの調査項目が設定され

ているが、今回は、本稿に必要なデータを転記した次の2種類の調査票を用いた。

- ・本人の職業経歴である職業データ部分のみを転記した「職歴票」
- ・本人の属性や父の職業データ等他の主な調査項目を転記した「コード・シート」

各調査票の説明に入る前に、1965年調査における職業データについて、1985年調査との違いを明らかにしておこう。1965年調査においても、1985年調査と基本的には同様で、注(5)に示す6種類のデータから構成されるが、大きく異なる点は、役名(1985年調査では役職に相当。以下、()内は1985年調査での項目名を示す)が自由回答であることである。このため、本来なら仕事の種類(仕事内容)で回答されるべきものが役名に出現する可能性が高く、分析時に注意する必要がある。

その他、名称やカテゴリーも若干異なっている。名称が異なるものは、仕事の種類(仕事内容)、役名(役職)、企業の規模(従業先の規模)であるが、本稿では、今後、混乱を避けるために1985年調査における名称で統一する。カテゴリーは、企業の規模は「10人未満、10人以上、100人以上、1000人以上」(1985年調査では、なし、1人以上、5人以上、30人以上、300人以上、500人以上、1000人以上、官公庁。以下同様)であり、従業上の地位は「主・役、単、家族、官雇用、民雇用」(同経営者(重役)・役員、常時雇用されている一般従業者、臨時雇用者・パート・アルバイト、自営業者(雇用者なし)、自営業者(雇用者あり)、家族従業者)である。これらはいずれも今回の分析に与える影響が大きくなり、変換も簡単に行えるため、特に問題とはしない。

各調査票の説明を行おう。まず、「職歴票」は、次の8種類(地点、個番、第何職目か、時期、事業、規模、役名、地位)の調査項目をもつ。以下、主なものについて簡単に説明する。

地点は地点番号で、01から50まであり、¹¹⁾それぞれ枝番として1、2をも

SSM職業データにおける自由回答の分析

つ。事業は職業データの従業先の名称および従業先の事業の種類を転記するための項目である。以下同様に、規模は従業先の規模、役名は仕事内容および役職、地位は従業上の地位をそれぞれ転記する項目である。

「コード・シート」は、本稿においては職歴票で不明な場合に参考にする程度で、それほど重要ではない。調査項目を注12) に示しておく。

1965年調査の東京地区におけるサンプル数は1252人で、各地点の平均は12人程度である。今回の分析は全部のデータではなく、地点番号01, 06, 11, . . . , 41, 45, (46の代用)¹³⁾のように、インターバルを5として抽出した10地点のデータを用いた。サンプル数は280人で、全体の約22%に相当するが、注11) から明らかなように、地域的に遍在している危険性はない。学生を除いて、サンプル1人が1個の職業経歴しか持たない場合が少ないために¹⁴⁾、本稿において分析の対象となる職業データのサンプル数は、最終的に957個であった。

2. 分析方法

分析は、自由回答である仕事内容と役職のデータを中心に、次の4項目について行った。

- ・ 仕事内容の形式的な分類
 - ・ 仕事内容の内容的な分類
 - ・ 仕事内容および役職で用いられた用語の頻度
 - ・ 上記用語と『SSM職業分類 (改訂版)』で用いられた用語との関連
- 以下、順に説明する。

(1) 仕事内容の形式的な分類

高橋 (1994b) で用いたサンプルでは、仕事内容のデータにいくつかのパターンがみられた (高橋1994c)¹⁵⁾。コンピュータは形式的に判断するため、最初に仕事内容における表現を形式的に分類することは意味があるものと思われる。役職については、形式的には同一のために特に行う必要は

ないとした。

(2) 仕事内容の内容的な分類

(1)での分類をさらに進めて、仕事内容のデータを内容面から分析する。これも、前述のサンプルにおいて、いくつかのパターンが見いだされたことによる（高橋1994c¹⁶⁾。ただし、ここでのパターンが異なっている限り、コンピュータはそれぞれを区別することができない。

(3) 仕事内容および役職で用いられた用語の頻度

辞書を作成する際の手語の選択に関して、直接、有効な情報となるものである。今回は特に、仕事内容だけでなく役職も自由回答であるために、同様の用語が双方に分散して出現する可能性があること、また、そもそも職業分類を判断する際に、管理的職業の判定には主として役職を用いることから、用語の出現頻度は両方の項目について調べることにした。

前述したように、1985年調査との対応は次のように行う。すなわち、最初は原データに従って分析を行う。次に、役職のデータについて、1985年調査であれば明らかに仕事内容または従業上の地位に回答されるであろうと予想できるものを適宜移動や削除により調整した後、再度、分析を行う。例えば、1985年調査における役職には選択肢「役職なし」があるが、これに該当する用語（一般事務員、店員、事務員等）は、同選択肢「その他」に出現するはずがないために、仕事内容に移動する。ただし、既に仕事内容に回答がある場合には、そちらを優先して削除し、ない場合にのみ実際の移動を行う。また、従業上の地位には選択肢「一般従業員」があるために、これが役職に出現するはずがないとして削除する。

自由回答を全体的に把握するには、送りがなの統一等多少の事前編集は行うが、基本的には用語をそのままの形でコンピュータに入力してソートを行う方法が有効である（高橋1995b）。本稿では、表計算ソフト（ロータス123）を用いて職歴票のデータを項目ごとに入力したファイルを作成し、

SSM職業データにおける自由回答の分析

五十音順に次の2種類のソートを行った。すなわち、キー1を仕事内容、キー2を役職とするソートおよびキー1を役職、キー2を仕事内容とするソートである。

前者では、仕事内容におけるデータの整理ができ、各用語の頻度が求まると同時に、仕事内容を同じ用語で回答したものについてすべての役職がわかる。同時に、後者では役職で用いられた各用語の頻度が求まり、同じ役職名をもつものについて仕事内容がすべてわかる。これにより、それぞれの回答においてどのような用語がどの程度用いられているかが明らかになり、さらに両者の回答における用語間の関連もわかる。もし、それぞれの回答または両者の回答において、用語間に類似関係や上位下位の階層構造を見いだすことができれば、単なる辞書¹⁷⁾ではなくシソーラスとできる。

(4) 上記用語と『SSM職業分類(改訂版)』で用いられた用語との関連

(1)から(3)までの分析は、『SSM職業分類(改訂版)』を意識しつつも、基本的には被調査により表現された回答の把握を目的とするものであった。ここでは、両者を積極的に関連付けた分析を行う。その際、両者を対等に置くよりも、どちらか一方に重点を置いて分析を行う方が効果的であると思われる。すなわち、次の2点について考察を行う。

- ・回答の『SSM職業分類(改訂版)』に対する関連付け
- ・『SSM職業分類(改訂版)』の回答に対する関連付け

特に、後者における結果は本稿における最終的な結論であり、辞書やデータベースの設計だけでなく、知識獲得の面でも有効になるものと期待できる。

Ⅲ. 結果と考察

全体としては、予想通り、前述したような自由回答の特徴がみられた。すなわち、表現のされ方が多様であったり、((2), (3)参照)、質問の意図に忠実な回答になっていなかったりした((2)参照)。以下、各分析項目ごとに

述べる。

(1) 仕事内容の形式的な分類

仕事内容を「の」、「で」、「を」等の助詞の有無により、形式的に分類した結果¹⁸⁾、次の4種類(A, BのA, CでA, DをA)があった。ここで、A, B, C, Dは助詞が含まれない部分を表す。各形式の頻度は、表1に示すように助詞がまったく含まれないA形式が圧倒的に多く、約70%を占める。これに、複数のA形式からなるA・A形式も含めると約80%近くになる。なお、表には示さないが、役職はすべてA形式であった。

A形式には、単一の単語からなる場合と、複数の単語からなる場合がある。例えば、「販売」は前者であるが、「機関車乗務員」は後者である。コンピュータは両者を区別できない点が問題である。

A・A形式は、「・」を取ることで簡単にA形式にすることができるが、Aが複数個あるときの優先順位をルールとして決めておく必要がある。一般に、各々を職業分類にコーディングしてからの方が決めやすいが、Aの段階で決めることができれば、処理時間の節約ができる。

BのA形式は、被調査者によってはA形式で回答されることもある。実際に、「パンの製造」と「パン製造」の両方の回答があった。コンピュータ

表1 仕事内容の形式的な分類

形式	例	頻度	割合(%)	
A	A	販売	686	71.7
	A・A	仕入・販売	61	6.4
BのA	花の小売 ¹⁹⁾	35	3.7	
CでA	デパート関係で工事 ²⁰⁾	3	0.3	
DをA	鉄をかつぐ	1	0.1	
無回答		171	17.9	
合計		957	100.0 ²¹⁾	

SSM職業データにおける自由回答の分析

処理の面からは、対象と動作が明確に分離しているBのA形式の方が処理しやすい。これは、調査員に対する事前の指導により可能になるものと思われる。

CでA形式、DをA形式はいずれも頻度が少ないために、当面の処理対象からは除外してもよいものと判断できる。

無回答は約20%近くあるが、役職ともに無回答のものは4%弱（34サンプル）であり、その場合もほとんどは従業先の名称が農業であるために、問題点にはならない。

(2) 仕事内容の内容的な分類

仕事内容はA形式（A・A形式を含む）が多かった。これを内容面から分類すると、A1(仕事の内容)、A2(職業名)、A3(職場名)、A4(扱う物品名)、A5(その他)の5種類があり、それぞれの頻度は表2のような結果になる。

A1がもっとも多く約70%弱、次がA2で約30%である。これらと比較すると少数ではあるが、職場名で回答される場合もあることに注意したい。これは、(3)で明らかのように事務部門が多かった。事務職の場合には、仕事内容を列挙するより職場名を回答する方が、被回答者と調査員の双方が理解しやすいとの判断であろうか。さらに少数になるが、扱う物品名によ

表2 仕事内容の内容的な分類(A形式)

内容	意味	例	頻度	割合(%)
A1	仕事の内容	販売	499	66.8
A2	職業名	販売員	221	29.6
A3	職場名	販売課 ²²⁾	16	2.1
A4	扱う物品名	ミシン ²³⁾	8	1.1
A5	その他	何でもやった	3	0.4
合計			747	100.0

る回答もあり、自由回答に対する分析のむづかしさを示す。この場合は、従業先の名称等他の項目を調べなければ判定することができない。

1985年調査の場合、約4万個の職業データをコーディングの対象としていることから、今回の結果がそのまま当てはまるものと仮定すると、1%程度の割合でも実際には無視できない個数となる²⁴⁾。従って、A3、A4に対する検討も行う必要がある²⁵⁾。

A1、A2の中には、同一の職業分類にコーディングできるものでも、表現の仕方が異なるものがある。一般に、職業名として認知されているようなものはA2で表現される場合が多い(表2のA1、A2の例を参照)。A2は形式的には、A1の後に、「員、人、者、士、師、家、屋、夫、工、労働者、社員、職、業務、仕事」が付く。「見習」が付く場合もある。A3は末尾が「部、課、係」である。これらを利用すれば、コンピュータに適切な指示を与えることが可能である。

これに対して、A1、A4はA2やA3とは区別できるが、末尾が不定なために相互の区別はできない。A5は、他に「転々として不明」、「一般」があるが、Aの1%未満しかなく、本稿では特に分析の対象としない。なお、表には示さないが、調整後に役職から移動された回答はすべてA2であった。

ここで、さらにA1を分類してみよう。基本的に、A11(Aのみ)、A13(内容的にBのA)、A14(職場名+A)の3種類に分類できるが、Aの末尾に「関係、一般、その他」が付くものをA12として区別した。頻度は表3の通りである。

一般に、内容的な区別はコンピュータには行えない。コンピュータができるのは形式的な分類で、これが可能なのはA12、A14だけである。いずれも個数が少なく、両方を合わせても5%程度であるが、サンプル全体の2.3%となるために、分析の対象とする必要がある。

ここでは、合わせて95%を占めるA11とA13を区別する方法について検

表3 A1の分類

内容	意味	例	頻度	割合(%)
A11	Aのみ	販売	361	72.3
A12	A+関係等	経理関係	10	2.0
A13	内容的にBのA	機械組立	116	23.2
A14	職場名+A	庶務課事務 ²⁶⁾	12	2.4
合計			499	100.0

討しよう。要は、内容をどのようにして形式に置き換えるかという問題である。次の2つの方法が考えられる。

第一の方法は、用語の先頭から機械的に2文字ずつ切り出していき、最後の語をAとし、それ以外をBとする。このとき、1語しか切り出せなければA11であり、その語はAとなる。例えば、「販売」であれば販売しか切り出せず、A11となって、販売がそのままAとなる。「機械組立」では、機械、組立と切り出し、前者をB、後者をAとする。「タービン部品製造」であれば、ター、ビン、部品がB、製造がAとなる。ここでの問題は、対象とする用語に1文字または3文字で一まとまりとなる語が含まれる場合である。²⁷⁾

第二の方法は、出現頻度の高い語をAのキーワードとして設定しておく、それを除いた部分をBとする。除いた部分がなければ、Aのみであり、A11となる。例えば、Aのキーワードとして「販売」を設定しておけば、販売がA11であるのは明らかである。「機械組立」では、Aのキーワードとして組立を設定しておけば、残りの語である機械がBとなる。「タービン部品製造」も製造をキーワードとしておけば、タービン部品がBとなる。第一の方法でうまくいかないものも、この方法ならできるが、対象とする用語がキーワードとして設定されていない語だけで構成されている場合にはできない。²⁸⁾

第一の方法を可能にするには、出現する可能性のある用語について、何文字で構成されているかについて調査する必要がある。第二の方法のためには、Aのみで出現する用語に決まったパターンがあるかどうか調査する必要がある。本稿では、後者の可能性を検討することとし、仕事内容がどのような用語で表現されているかについて調査した。その結果を(3)に述べる。

(3) 仕事内容および役職で用いられた用語の頻度

分析の方法で述べたように、両項目とも原データと調整後のデータについて調査する必要があるが、紙面の制約上、原データを中心に主なものをそれぞれ表4, 5に示す。送りがな以外は、被調査者により表現された用語をそのまま用いた。

・仕事内容

仕事内容においては、Bを表現する用語は多岐にわたるため、一部を注29)に示すにとどめ、ここでは、Aを表現する用語を中心に、主として複数個出現したものを示した。表中、()のあるものは、最初に示した用語の他にそれに続く語をもつ用語があることを意味する。複数ある場合は、「, 」で続けた。その際、頻度はそれぞれを「+」で結んで表す。また、意味だけでなく字句においても類似の用語があるときは、「, 」で続けた。合計は、1985年調査との調整により、役職より移動されるものも含めた数値を示す。

表4で示した用語の頻度を合計すると約47% (454) 強であるが、学生が約1%弱 (8), 無回答が約20%弱あるために、これに基づいてキーワードを設定した場合に切り出せない用語は、単純に計算すると約32%である。しかし、表4は用語の先頭文字によるソート結果をまとめたものでしかなく、実際のシステム稼働時には、キーワードが用語の途中や最後にくるものも切り出すことが可能である。例えば、製造をキーワードとすると、パン製造や紙製品製造等も切り出せる。また、今回、Bについては考慮して

表4 仕事内容に出現した主な用語

用語	頻度	合計
セールス(マン)	4+1	5
プレス型加工, プレス工	1,8	9
一般事務(員)	12+1	15
一般店員	1	1
医師, 医員	5,1	8
運転(手)	2+18	20
運搬(さしず, 従業員)	5+1+1	7
営業(一般, 事務, 課, 部事務)	9+1+1+2	13
卸(売, ・販売)	1+2+1	4
加工(仕上, ・販売)	3+2+1	6
家事見習, 家事(の)手伝	1,4+1	6
外交(社員, ・経理, ・仕入・販売, ・洗濯, 配達)	2+1+1+1+2	7
管理	4	4
企画	2	2
機械運転士, 機械操作, 機械監視	1,1,1	3
機械整備, 機械修理	1,1	2
機械助手見習	1	1
機械, 機械組立(工)	2,1+1	4
教員, 教師, 教諭	21,4,5	30
経営(・営業部監督, ・鉄工の仕事, ・販売手伝)	1+1+1+1	4
経理(関係, 事務, 部, 部会計課)	13+4+6+1+1	25
建築, 建築資材の係	4,1	5
見習(工, 店員)	6+1+1	9
工員	2	2
行商	2	2
採炭(夫)	1+1	2
雑役(夫)	2+1	3
雑用	2	2

用語	頻度	合計
仕上(工, 検査, ・組立, 過程監督)	4+4+1+2+1	12
仕入(販売, ・販売, ・運搬・販売・配達)	3+1+3+1	8
事務(員, ・販売, ・加工, 管理)	22+2+2+1+1	31
車掌	4	4
手伝	7	7
出前(持)	2+1	3
庶務(一般, 課事務)	1+1+1	3
小売販売	1	1
整理	2	2
製造(販売, ・卸, ・販売, 手伝, 過程)	8+1+1+4+2+1	17
設計・研究, 設計(事務, ・事務)	1,2+1	4
旋盤((職)工), ・機械工	7+8+1	16
組立(工(員)), ・仕上	7+4+4	15
大工(・現場見回, ・棟梁, 見習)	8+1+1+2	12
調理(師, 人, ・接客)	2+1+1+1	5
電気(器具組立, 工事, 修理)	1+1+1	3
店で販売, 店員	4	5
塗装(工)	5+1	6
土木技術, 土木建設, 土木工事, 土木職	3+2+1+1	7
農業(・製紙加工), 農作業, 農仕事(手伝)	1+1+1+3+1	7
農作物の売買	1	1
配達(店員, ・セールス・事務, ・外交, ・製造, ・販売)	2+2+1+1+1+1	8
販売(員, 店員, 事務員, 見習, 手伝, その他, 部事務, 係, ・配達, ・集金, ・整理, ・行商)	39+1+5+1+1+1+1+1+1+1+1+1+1	55
部品(加工, 製造, 組立, 分配)	1+1+1+1	4
溶接(工)	4+3	7
料理	2	2
露天商	2	2

SSM職業データにおける自由回答の分析

いないが、これについてもキーワードを設定すれば、切り出せない用語の割合はかなり低くなるはずである。

用語を切り出すキーワードの候補として、表4から合計が10以上を基準に選ぶと、一般事務、運転、営業、教員、経理、仕上、事務、製造、旋盤、組立、大工、販売がある。

このうち、製造および販売を取り上げて、具体的な用例を示す。

製造が先頭にあるもの

製造 製造販売 製造・卸 製造・販売 製造手伝 製造過程

製造が最後にあるもの

コンニャク製造卸 タービン部品製造 紙製品製造 蛇口製造 電球製造
建築材製造 魚雷製造 部品製造 パン製造
ネジなどの製造 電球の製造 パンの製造

販売が先頭にあるもの

販売 販売員 販売店員 販売事務員 販売見習 販売手伝 販売その他
販売部事務 販売係 販売・配達 販売・集金 販売・整理 販売
・行商

販売が最後にあるもの

卸・販売 医薬品販売 書籍販売 肥料販売 靴の仕入・販売

これらの語が最後にあるものは、いずれも（内容的な分類も含めて）BのA形式である。一般に、製造の場合は、職業大分類³⁰⁾では熟練的職業または半熟練的職業等に判定されることが多いため、Bの違いにより職業分類が異なるが、販売の場合は販売的職業になるために、Bは関係せず、役職が経営主または店主であるかどうか、もしそうであればAが販売以外に卸を含むかどうかで決まる³¹⁾。従って、製造ではBの語が重要であり、販売では役職が重要になる。

この意味で、上記の例では仕上、組立が製造と、営業、経理、事務が販売と同様である。教員は他の語が付く場合には必ずその用語の末尾にあり、先頭の語の違いにより職業小分類が異なる。例えば、小学校教員、中学校教員、高等学校教員、大学教員は、それぞれがそのまま異なる職業分類名である。³²⁾

実は、キーワードとして設定した語により注目する部分や質問項目が異なることは、『SSM職業分類(改訂版)』からの知識であるが、実際にどのような用語で回答されるのかという実態を把握しなければ、活用することができなかった。

以上より、仕事内容においてはキーワードとする用語の種類により、注目する項目や整理する用語が異なることが明らかになった。

・役職

前述したように、役職においても用語の混乱がみられる。原データの主なものを表5にまとめる。大まかに職業大分類別に分け、1985年調査であれば明らかに役職には回答されないと予想できるものをグループの下部に記した。その結果、依然として役職に残る用語の種類はそれほど多くない。³³⁾ なお、無回答は291(約30%)で仕事内容より多いが、質問の性質上、当然であろう。

役職においては、表5に示したように役職名だけからなる用語だけでなく、以下に示すように、職場名+役職名による回答がみられる。その割合は、無回答を除いた原データの約6%(40)である。課長より下位のもの管理職とはならないために、職業分類を判断する際には仕事の内容を暗示する職場名が重要になるが、これは役職名をキーワードとすれば簡単に切り出せる。

部長 市内部長 事務部長 受け渡し部長 総務部長 販売部長

次長 運動部次長 経理次長 工場次長 整理部次長

課長 運用課長 仕上課長 受付課長 庶務課長 整備課長 第一料金

表5 役職に出現した主な用語

用語	頻度	用語	頻度
(半, 非)熟練的職業		専門的職業	
職長	7	技官	3
班長	13	技師	3
組長	5	研究員	4
		研究主任	1
見習工	5	訓導	5
工員	33	校長	9
一般工	41	助手	5
一般工員	13	講師	2
並工(員)	2+3	助教授	2
作業員	3	教授	6
一般作業員	2	病院長	1
工手	2		
技手	2	教官	1
技手補	1	教師	1
職工	6	教諭	4
職人	2		
理容師	1	事務的職業	
採炭夫	1	主事(補)	5+1
雑役夫	1		
人夫	3	事務	5
コック	1	一般事務(職)	5+2
		事務員	6
		一般事務員	11
		事務官	4
		事務職員	3

用語	頻度
販売的職業	
店主	43
支配人	5
店員	38
一般店員	3
管理的職業	
課長	11
次長	4
部長	5
工場長	5
(出張)所長	1+3
(常務)取締役	2+4
専務(取締役)	2+2
相談役	1
理事	1
重役	5
社長	4
参与	1
経営主	40
経主	3
経営者	4
村助役	1

用語	頻度
特定できないもの	
主任	7
係長	17
課長代理	3
嘱託	4
従業員	3
一般従業員	22
社員	3
職員	6
一般職	3
一般職員	6
所員	2
小僧	10
見習	10
技術員	7+1

課長 調査課長 統計課長 秘書課長 保管課長 料金経理課長
 係長 会計係長 外務係長 技術課研究係長 仕上係長 施行係長
 事務係長 出納係長 整備係長 団体係長 総務係長 徴収係長
 主任 運搬主任 各階主任 指導主任 編集主任 倉庫課主任 鋳物主

任 溶接主任
室長 検査室長
主事 厚生係主事

・仕事内容と役職の関係

最後に、仕事内容と役職の関係について簡単に述べておく。同一の仕事内容に対して同一の役職名があるものより、役職なしに該当するものから課長以上のものまで、さまざまな役職名が出現するものの方が多い。セールス、一般事務、運転手、運搬、営業、卸、加工、外交、経理、建築、仕上、仕入、事務、庶務、製造、旋盤、組立、大工、塗装、販売、溶接において、この傾向がみられた。

SSM職業コーディングにおいては、生産第一主義を採用する立場上³⁴⁾、できるだけ管理的職業に判定しない方針をとる。従って、上記の中にも管理的職業に判定される役職名をもつものも含まれるが、仕事内容がいずれも管理や経営以外であるために、こちらを採用しなければならない。

これより、職業分類の判定に際しては、最初に仕事内容を調べることで、その結果が役職より得られた結果と異なる場合には、仕事内容による結果を優先することの2つのルールが必要であることがわかる。

(4) 上記用語と『SSM職業分類(改訂版)』で用いられた用語との関連
・回答の『SSM職業分類(改訂版)』に対する関連付け

ここでは、これまでに得られた回答の分析結果が、どの程度『SSM職業分類(改訂版)』で利用できるかについての考察を行う。具体的には、回答に出現する用語が、『SSM職業分類(改訂版)』にどの程度出現するか、言い替えれば、両者で用語のレベルが合っているかという問題として設定できるが、これを仕事内容と役職(調整後のデータ)に分けて述べる。

最初に、仕事内容について述べる。表4に示された用語は主としてAであったが、このうち、『SSM職業分類(改訂版)』に存在するのは、52個中

48個で約92%である。存在しなかったのは出前，農業，農作物，整理の4個だけで，予想に反して結果がよい。ただし，これがそのまま職業分類の判定につながるわけではなく，例えば，見習いは地位しか表さないし，加工，仕上，製造，組立も「何を」を表す用語が不明なために決定できない。一方，この用語だけで職業分類が決定できるのは，セールス（「071 外交員（保険，不動産を除く）」の職業例にある），一般事務員（「055 一般事務員」），医師（「013 医師」）等で，約62%（32個）である。

決定度を向上させるには，次の2つの対策が考えられる。1つは，例えば，農作物を「穀物，野菜，果樹」（注（10）「076 農耕・養蚕作業」参照），出前を配達とみなすことができるように，上位下位関係や類似関係にある用語をあらかじめコンピュータに教えておく。すなわち，シソーラスを作成しておく。

もう1つは，前述した加工，仕上，製造等キーワードとできるような用語については，各用語別に「何を」に相当する用語を詳細にリストアップしておき，それにより決定される職業分類と対にしてコンピュータに教えておく。このためには，キーワードとできるAの用語については，ともに出現するBの用語について整理しておく必要がある。これについては，次の分析項目で具体的に検討する。

役職における回答の用語は，表5より明らかなようにある程度限定されている。ここでは，職業分類の判定に用いるような用語が重要であるが，これはさらに限られた範囲内のものしかなく，多くは一般的な用語である（例えば，社長，部長，課長等）。『SSM職業分類（改訂版）』では一般的な用語が用いられるため，両者の一致度は高くなり，コンピュータによる処理が行いやすい。

・『SSM職業分類（改訂版）』の回答に対する関連付け

最後に，これまでの結果に基づき，回答を解釈するためには，『SSM職業分類（改訂版）』の内容をどのように整理すればよいかについて考察す

SSM職業データにおける自由回答の分析

る。具体的には、回答から職業分類を判定する際にキーワードとする用語Aの『SSM職業分類（改訂版）』における用いられ方を調査し、各々のAに関係するBの用語を整理する。

Bの有効性を職業大分類別にみると、前述したように、（半・非）熟練的職業で最も高く、事務的職業、販売的職業、管理的職業は比較的低い。専門的職業は中間である。ここでは、紙面の都合上、専門的職業について考察した結果を示す。

表6は、表4に示した用語のうち、専門的職業に分類できるものについて整理したものである。表中、 \approx は類似関係、 \rightarrow は上位下位関係（矢印の方向が上位）を表す。『SSM職業分類（改訂版）』欄は職業小分類名およびそのコードを示す。左詰めのもはAの用語のみかA、Bの用語により判定できることを表し、右詰めのもは従業先の名称や事業内容も判定に必要なことを表す。ここには、A、Bだけで判断できない場合に従業先データをみるというルールが適用される。

専門的職業は48個の職業小分類（001～045，101～102，108）からなるが、このうち13個のものが表6に示される。

表6で得られた知識を基にして、『SSM職業分類（改訂版）』から獲得できる知識を表7にまとめる。表中、A、Bの用語は、『SSM職業分類（改訂版）』において用いられる用語である。備考欄において「Bの用語を上位下位、類似関係で整理が必要」というのは、『SSM職業分類（改訂版）』での記述や一般的な用例に従ってBの用語を上位とする用語をできるかぎり多く収集して下位の用語とする階層構造を構築すること（上位下位関係で整理）、およびBの用語と同様の意味で用いられる用語を収集すること（類似関係で整理）をいう。例えば、自然科学系の下位用語は、注17)に示すような物理学、数学を始め、統計数理、気象、地質、天文等³⁵⁾があり、さらにこれらの用語を上位とする下位用語（例えば、物理学に対して核融合、プラズマ等）が収集できる。

表6 表4における専門的職業の整理

Aの用語	Bの用語	『SSM職業分類(改訂版)』	従業先
研究	核融合・プラズマ	→ 自然科学系(001)	研究所
	なし	機械(005) ←	農器具
	技術	電気(006) ←	高圧容器製造
	製鉄	→ 金属精錬(004)	
	高圧容器	→ 電気(006)	
	鋳物	→ 金属精錬(004)	
技術	X線	その他の保険医療(021)	
〃	建設	建設≈土木(009)	
技師	建設局	同上(009)	
	受信管標準化	→ 電気(006)	
	電気	(006)	
	土木	(009)	
医師	なし	(013)	
〃			
医員			
弁護士	なし	(022)	
教員	なし	小学校教員(026)	小学校
〃		小学校教員(026)	**小学校
教師	なし	中学校教員(027)	中学校
〃		中学校教員(027)	中学
教諭		高等学校教員(028)	高等学校 高等実業学校 青年学校
なし	なし	大学教員(029) (ただし役職が教授, 助教授, 講師, 助手)	**大学
記者	なし	(034)	
編集	なし	(034)	
(誤った例)			
技術	映写	映写技師は(254)にあり	
〃	関係	(半・非)熟練的職業(041～)	
技師		(役職が工員)	

表7 『SSM職業分類(改訂版)』から獲得される知識(専門的職業の場合)

Aの用語	Bの用語	職業小分類コード	備考
研究	自然科学系	001	Bの用語を上位下位, 類似関係で整理が必要。 従業先が研究所でなければ003~012。 Bの用語を上位下位, 類似関係で整理が必要。
	人文科学系	002	
技術	鉱山	003	
〃	金属精錬	004	
技師	機械	005	
	電気	006	
	化学	007	
	建築	008	
	土木	009	
	農林	010	
	情報処理	011	
	無線通信	108	
	その他	012	
医師	なし	013	Aのみで判定可
歯科医	なし	014	Aのみで判定可
薬剤師	なし	015	Aのみで判定可
助産婦	なし	016	Aに見習がないことを確認。 見習ありは021(017, 019~020でも同様)。
保健婦	なし	017	Aに見習がないことを確認
栄養士	なし	018	Aのみで判定可
看護婦	なし	019	Aに見習がないことを確認
看護師	なし	019	Aに見習がないことを確認
あん摩	なし	020	Aに見習がないことを確認
はり	なし	020	Aに見習がないことを確認
きゅう	なし	020	Aに見習がないことを確認
柔道整復	なし	020	Aに見習がないことを確認
裁判官	なし	022	Aのみで判定可
検察官	なし	022	Aのみで判定可
弁護士	なし	022	Aのみで判定可
公認会計士	なし	024	Aのみで判定可

Aの用語	Bの用語	職業小分類コード	備考
税理士	なし	024	Aのみで判定可
教員		025～031	従業先で区別(表6参照)
〃			(幼稚園, 小学校, 中学校, 高等学校, 大学,
教師			盲学校・ろう学校・養護学校, その他の学
〃			校)
教諭		044	従業先がないもの
宗教	なし	032	Aのみで判定可
文芸	なし	033	Aのみで判定可
著述	なし	033	Aのみで判定可
記者	なし	034	Aのみで判定可
編集	なし	034	Aのみで判定可
彫刻	なし	035	Aのみで判定可
画家	なし	035	Aのみで判定可
工芸美術	なし	035	Aのみで判定可
デザイナー	なし	036	Aのみで判定可
写真	なし	037	Aのみで判定可
カメラマン	なし	037	Aのみで判定可
音楽	なし	038	Aのみで判定可
俳優	なし	039	Aのみで判定可
舞踏	なし	039	Aのみで判定可
演芸	なし	039	Aのみで判定可
プロ	なし	040	プロの後の語がスポーツ名
獣医	なし	041	Aのみで判定可
保母	なし	042	Aのみで判定可
保父	なし	042	Aのみで判定可
福祉	なし	043	Aに事務がないことを確認
機関	船舶	101	Bの用語を上位下位, 類似関係で整理が必要。
	航空機	102	
操縦	航空機	102	Bの用語を上位下位, 類似関係で整理が必要。
航空	なし	102	Aのみで判定可
通信	無線, 船	108	
	舶, 航空		

表7より明らかなように、専門的職業における知識は比較的単純であり、Aの用語だけで判定できるものも多い。また、前述したA2形式の末尾にくる可能性のある語は「師、士、家」で、「屋、夫、工、労働者」はないことがわかる。

以上、ここでは専門的職業について述べたが、他の職業大分類においても、同様の作業を行えば知識獲得が可能である。

おわりに

本稿では、1965年SSM調査における一連の職業データのうち、主として自由回答である仕事内容や役職において表現される用語の分析を行った。これは、職業データをコンピュータで解釈しようとする「SSM職業コーディング支援エキスパート・システム」における辞書やデータベースの構築に必要であるが、知識ベースを構築するための知識獲得にも有効であった。

本稿での結論をまとめると、次の通りである。

- ・回答は大部分がA形式またはBのA形式であるが、内容的にはいくつかのパターンがある（コンピュータは内容は区別できない）。
- ・被調査者により表現された回答と、データベースや知識ベースの源となる『SSM職業分類（改訂版）』における用語は、抽象度においてそれほどの違いがない。
- ・ただし、仕事内容においては、前者の用語は多様であり、後者のものと上位下位関係や類似関係にあるものが出現する。上位下位関係はB、類似関係はAに多い傾向がある。
- ・職業分類を判定する際に、まず、職業大分類の違いにより注目する項目や語が異なる。さらに同じ職業大分類でも、職業小分類の違いにより同様のことがいえる。

- ・本システムにおいて効率のよい辞書とは、全体を均一にするのではなく、必要な部分のみ密にすることを意味する。これは、知識獲得においても同様である。

結論の中には、単に用語の対応関係だけでなく、ルールとして作成できる知識も含まれる。今後の課題は、本稿での結論を活用した辞書やデータベース、さらには知識ベースの構築を行うことである。

謝辞

『社会調査演習』に掲載された職業データを職業小分類にコーディングして下さった原 純輔東北大学教授、並びに1965年SSM調査の職歴票とコード・シートを送付して下さった佐藤俊樹東京工業大学助教授に深く感謝いたします。さらにこれを許可して下さった1995年SSM調査委員会の皆様に感謝いたします。

- 注 1) 1955年に第1回目の調査が実施されて以来、10年ごとに行われている。
- 2) 原(1986)によれば、職業の変化は次の4つの場合をいう。
- ・従業先の変化(自営業者の商売替えを含む)
 - ・従業上の地位/役職の変化
 - ・仕事の変化
 - ・解雇、引退、兵役、就学などによる職業生活の中断
- 3) より正確には、1955年調査(第一回)からの継続項目は、本人の現職、職業経歴、父の主な職業、本人就職時の父の職業であり、1965年調査では配偶者やその父親は含まれない。しかし1985年調査では本文で述べたものに加えて、本人が15才時の父職も調査項目とされ、1サンプルにつき平均10個の職業をコーディングする必要があった(原1993)。
- 4) これは、国勢調査でも同様である。
- 5) 職業データは、3種類の自由回答(従業先の名称、従業先事業の種類、本人の仕事内容)と3種類の選択回答(従業上の地位、従業先の規模、役職名)から構成される。
- 6) 例えば、1985年調査の場合、対象となる職業データは約4万個(3947サンプル×10個)あり、20~30人のメンバーの泊まり込み作業でほぼ1週間に要した(原1993)。

SSM職業データにおける自由回答の分析

- 7) 一般に、エキスパート・システムで扱うデータは数値やコードのような簡単な情報が多い。しかし、注5)で明らかのように、職業データのすべてをコード(選択回答)にすることはできない。
- 8) 極端な例としては、まったく同じ用語でも、表記が異なれば別のものとみなされる。例えば、ペンキとぺんき、見習いと見習等。
- 9) 高橋(1994b)では、職業小分類(293種類)にまでは分類せず、職業大分類(8種類)にとどまっていた。今回の結果については、資料を参照のこと。
- 10) 「076 農耕・養蚕作業」の職業定義文および職業例は以下の通りである。

穀物、野菜、果樹その他の作物の栽培、収穫などの作業及び蚕の飼育、収繭、蚕種の製造などの作業に従事するものをいう。

○農耕作業者、○園芸農耕者、○稲作農耕者、○野菜栽培者、○茶栽培者、○きのこ類栽培者、○しいたけ種駒植付人、○もやし栽培作業、○草花栽培者、○造林用苗木栽培者、○芝栽培者、○養蚕作業、○蚕飼育者、○蚕種製造作業、○農耕実習助手、○耕作用トラクタ運転手
×開墾作業 [228]、×製糸工 [149]、×もやし製造工(工場で行うもの) [216]、×盆栽作り人 [083]

- 11) 地点番号と地点名の対応は、次の通りである。

地点番号	地点名	地点番号	地点名
01, 02	練馬区	27, 28	北区
03, 04	杉並区	29, 30	板橋区
05	不明	31	足立区
06	中野区	32	葛飾区
07, 08	不明	33, 34	江戸川区
09, 10	世田谷区	35	葛飾区
11	不明	36, 37	足立区
12	大田区	38	荒川区
13	不明	39	不明
14	目黒区	40	江東区
15	千代田区	41, 42	墨田区
16	港区	43, 44	台東区
17	不明	45	中央区
18, 19	新宿区	46	港区
20, 21	不明	47	品川区
22	渋谷区	48	不明
23	不明	49, 50	大田区
24	豊島区		

25, 26 不明

- 12) コード・シートの調査項目は、次の通りである。
地点, 個番, 出生年, 出生地, 東京生まれか, 転入者か, 転入期 (転入者の場合), 親と一緒に (同), 父出生地, 最終学校, 学校地, 卒業年, 学校歴, 父学歴, 諸職時の父の世話, 父事業, 父職規模, 父職役名, 父職地位, 余暇, 個人収入, 投票正当, 支持政党, 住居
- 13) 地点46(サンプル数22)のデータが一部不鮮明であったため, 45(サンプル数31)で代用した。
- 14) 調査時期の関係上, 被調査者の年齢によっては, 職業生活が戦争により中断される場合があるために, 他の時期に比較すると職業の変化が多い傾向がある。
- 15) 例えば, 何々といいきる形 (販売員, 広告計画立案等), 何々の何々という形 (工場の管理全般, 眼鏡の調整等), 何々と何々という形 (金属プレスと外交等), 何々・(または,) 何々(事務・販売, 販売, 修理等) という形があった。
- 16) 例えば仕事の内容を説明するもの (経理, 米作等), 職業名 (販売員, 警察官等), 職場名 (総務部人事課, ベビー服売場等) があった。
- 17) 本稿の目的からいえば, 例えば教員, 教諭, 教師は類似語とみなすことができる。また, 自然科学と物理学や数学は上位下位の階層関係にある。
- 18) コンピュータ・システムにおいては, 文字は各々対応するコードが定められているが, どのコード体系においても記号, ひらがな, カタカナ, 漢字はそれぞれまとまった範囲内にある。例えば, パーソナルコンピュータで一般的なアスキー (ASCII) コードの場合, 記号は8120~8140(16進数表記による。以下同様), ひらかなは829F~82F1, カタカナは8340~8396, 漢字は889F~の範囲にある。
- 19) BのA形式は, 他に次のものがあった。
- | | |
|-------------|----------|
| すしのにぎり | 工場の電車の電気 |
| タタミ表の検査 | 諸々の手伝 |
| ネジなどの製造 | 嘱託の司法事務 |
| ソケットの製造 | 船の燃料購入 |
| 電球の小売 | 製品の発注 |
| レンズの像透 | 総務の事務 |
| 家庭の手伝 | 都電の車掌 |
| 核融合・プラズマの研究 | 農作物の売買 |
| 魚の運搬・配達 | 配給の仕事 |
| 魚の出し入れ・運搬 | 売店の経理事務 |
| 魚の身をつぶす | 募集員の監督 |
| 靴の仕入・販売 | 募集員の指導 |

SSM職業データにおける自由回答の分析

- 鉄工の仕事
建築資材の係
- 無線の計画・施工・保守
毛の洗浄
- 20) CでA形式は、他に次のものがあった。
食堂で雑役
店で販売
- 21) 各形式ごとにまとめを行うため、合計が100%にならないことがある（以下の表も同様）。
- 22) 他に次のものがある。
- | | | | |
|--------|-----|-------|------|
| 本社事業部 | 営業課 | 為替係 | 鋳物現場 |
| 経理部 | 工作課 | 元帳係 | |
| 経理部会計課 | | 出納係 | |
| | | 財務相談係 | |
| | | 預金係 | |
| | | 工事係 | |
| | | 倉庫係 | |
| | | 市電検査係 | |
- 23) 他にアイロン（ミシン・アイロン）、ろくろ、巻線、鋳物、通信機器部品、電気一般、細菌学がある。
- 24) 例えば、今回のデータに占める割合が1%でも400個になる。
- 25) 全体に対する割合は、それぞれ16.7%、0.8%である。
- 26) 他に、次のものがある。
- | | | |
|-------|----------|-------|
| 貸付係書記 | 営業部事務 | 営業部監督 |
| 旅客係書記 | 販売部事務 | 建設局技師 |
| | 地方課事務 | 社会部記者 |
| | 財務課事務 | 統計課員 |
| | 商品課事務 | |
| | 違法監督係事務員 | |
- 27) 例えば、医薬品製造、建築材設計、機械試運転という用語がある。
- 28) 例えば、電報授受等の特殊な用語は、キーワードとして設定される可能性が低い。
- 29) 例えば、次のようなものがある。
- かばん+制作, 加工, 裁断・材料仕入 ガラス+工事, 拭工 コンニャク製造卸 セメント+混ぜ, 埋め合わせ タービン部品製造 タタミ表の検査 ネジなどの製造 バケツ制作 パンの製造 パン製造 部品+加工, 製造, 組立(工), 分配
- 30) SSM調査における職業大分類は、専門的職業、管理的職業、事務的職業、販売的職業、熟練的職業、半熟練的職業、非熟練的職業、農林的職業の8分類をいう。
- 31) 役職が経営主または店主であり、かつAに卸が含まれれば卸売店主、含

- まれなければ小売店主である。
- 32) その他に，盲学校・ろう学校・養護学校教員，その他の教員がある。
- 33) さらに，職業分類を判定する際に，役職に残った用語のすべてが用いられるわけではない。
- 34) この立場によると，各職業大分類の優先順位は次の通りである（くは左辺の方が右辺より優先度が低いことを表す）。
管理的く事務的く販売的く（半・非）熟練的，農林的
- 35) これらは『SSM職業分類（改訂版）』で用いられた用語の一部である。

参考文献

- 原 純輔 1986.「職業移動のネットワーク」『リーディングス日本の社会学8 社会階層・社会移動』（直井優・原純輔・小林甫編）：214-228 東京大学出版会
- 1993.『SSM職業分類（改訂版）』（文部省科学研究費補助金研究結果報告書）
- 海野道郎 1984.『社会調査演習』東京大学出版会
- 高橋和子 1994a.「職業分類における自然言語処理について——知識ベース・システム構築のための予備的考察——」『千葉敬愛短期大学紀要』 Vol.16: 127-140
- 1994b.「社会調査におけるエキスパート・システム構築について——SSM職業コーディング支援エキスパート・システムの構想——」『国際教養学論集』 Vol.4: 143-172
- 1994c.「SSM職業コーディング支援エキスパート・システムの構築について」『第18回数理社会学会大会研究報告要旨集』：57-58
- 1995a.「社会調査におけるエキスパート・システム構築について——SSM職業コーディング支援エキスパート・システムの構想——」『情報処理学会第50回全国大会講演論文集（1）』：323-324
- 1995b.「非定型データの処理・分析について」『札幌学院大学社会情報学部紀要』 Vol.4 No.2: 107-121
- 野崎 進他 1993.「アンケートにおける日本語自由文の情報分析」『情報処理学会第47回全国大会講演論文集（3）』：165-166

第3回 SSM調査における職業データのコーディング結果例

(データは原・海野(1984)による)

番号	年齢 (歳)	最終 学歴	従業先		規模*	職業			職業 小分類 コード
			名称	事業内容		地位	仕事内容	役職	
001	55	中学	〇〇農協	ガソリン販売	3	雇用者	販売員	067	
002	38	高校	〇〇精機	カメラ部品製造	4	雇用者	経理	056	
003	67	不明		農業	1	自営	米作	076	
004	28	高校	〇〇警察署	〇〇駐在所	8	雇用者	警察官	264	
005	41	中学	〇〇食品	〇〇	3	雇用者	工場の管理全般	051	
006	27	大学	〇〇インフイルコ	下水処理のプラント設計	7	雇用者	総務部人事課	055	
007	37	中学	〇〇製鉄	製鉄業	4	雇用者	精選工	120	
008	45	大学	〇〇工機	工業機械の販売	5	雇用者	企画	051	
009	25	高校	〇〇青果	果物の委託販売	5	雇用者	事務・販売	067	
010	52	高校	〇〇県庁		8	雇用者	一般事務	055	
011	50	高校	〇〇織物	裏絹製造販売	2	自営	いっさい	154	
012	34	高校	〇〇サイクル	自転車販売	1	自営	販売・修理	142	
013	42	大学	〇〇光学機械K.K	光学精密機械の製造	7	雇用者	経理部主計課	051	
014	28	高校	レストラン	飲食店	3	雇用者	ウエイター	278	
015	41	中学	国鉄	鉄道業	7	雇用者	貨物列車運転	097	
016	41	中学	〇〇ポンプ	ポンプ製造	5	雇用者	旋盤工	121	
017	38	中学	〇〇製作所	モーターのケース(鋳物)	7	雇用者	造形工(鋳型)	115	
018	33	高校	〇〇製作所	鉄加工	4	雇用者	機械加工	121	
019	43	高校	〇〇商工リサーチ	興信所	7	雇用者	庶務	115	
020	24	大学	〇〇ストア	スーパーストア	5	雇用者	ベビー服売場	067	
021	39	大学	〇〇建設工業	電話線の架設	7	雇用者	電話線工事	236	
022	46	高校	〇〇食店	中華料理店	3	自営	コック	276	
023	56	大学	〇〇中学校	中学校	8	雇用者	校長	027	
024	50	大学	〇〇商店	日用品・食料品の販売	1	自営	仕入 小売 配達	064	

番号	年齢 (歳)	最終 学歴	従業先			職業			職業 小分類 コード
			名称	事業内容	規模*	地位	仕事内容	役職	
025	43	中学	〇〇産業	砂利運搬 販売	4	雇用者	ダンクパーの運転	099	
026	44	大学	〇〇信用金庫	金融業	7	雇用者	電子計算機係	063	
027	43	中学	〇〇石炭〇〇鉱業所	石炭採掘業	7	雇用者	採炭員	090	
028	30	中学	〇〇理容室	理容業	3	雇用者	理髪師	272	
029	54	中学	〇〇病院	精神病院	4	雇用者	調理師	276	
030	53	中学	〇〇研究室	建設用鉄筋の強度検査	3	雇用者	検査	120	
031	33	高校	〇〇電器	コンデンサ製造	6	雇用者	製造	051	
032	45	中学		漁業	1	自営	漁師	084	
033	52	高校	〇〇〇〇コングリート工業	生こん製造	3	雇用者	料理	056	
034	62	中学	〇〇洋服店	洋服仕立販売	1	自営	裁断	160	
035	51	中学		稲蘭草	1	自営	稲蘭草	076	
036	40	中学	〇〇石材店	墓・石碑造	3	重役	塑刻	201	
037	55	高校	〇〇宮林署	国有林の管理	8	雇用者	庶務	046	
038	22	大学	〇〇眼鏡店	眼鏡の販売	2	家族	眼鏡の調整	148	
039	24	高校	〇〇コンピュータシステム	データ調べ	4	雇用者	プログラマー	011	
040	25	高校	〇〇自動車工業所	自動車の板金塗装	3	雇用者	塗装	241	
041	25	大学	株式会社	広告業	7	雇用者	広告計画立案	055	
042	30	大学	〇〇社東京本社	出版業	5	雇用者	編集	034	
043	43	中学		農業(りんご, 大豆, 米)	1	家族	農業	076	
044	61	高校	〇〇商事	サービスマン	4	雇用者	食堂などの雑役	287	
045	32	中学	〇〇金属	金属プレス	3	雇用者	金属プレスと外交	122	
046	22	大学	〇〇水上警察署		8	雇用者	英語通訳	045	
047	51	高校	〇〇装置	プラント工事組立	4	雇用者	企画	008	
048	54	大学	〇〇〇〇〇センター	雑役下請業	3	重役	経営	051	
049	36	高校	〇〇水産船舶部	漁業および水産物加工	7	雇用者	二等航海士	085	
050	31	中学	〇〇工務店	大工	2	自営	大工	221	

SSM職業データにおける自由回答の分析

番号	年齢 (歳)	最終 学歴	従業先			職業			職業 小分類 コード
			名称	事業内容	規模*	地位	仕事内容	役職	
051	49	高校	〇〇〇区役所		8	雇用者	公園建設統括	建設係長	055
052	66	中学	〇〇〇商店	菓子, 雑貨販売	1	自営	販売		064
053	27	高校	〇〇〇計器	ガス器具等の販売	4	雇用者	サービスマン	主任	067
054	34	大学	電電公社	電報・電話	7	雇用者	架設工事技師	主事	006
055	32	中学	〇〇〇学園高校		5	雇用者	調理師		276
056	34	大学	〇〇〇大学化学研究所	大学	8	雇用者	抗癌物質の開発合成	助手	029
057	56	大学	〇〇〇港管理事務所	港湾管理 施設貸与	8	雇用者	野球場の使用許可など		055
058	46	中学	米穀卸商業協同組合	米の販売	3	雇用者	米の加工		203
059	58	中学	〇〇〇家具店	家具小売店	2	雇用者	店員		067
060	47	高校	〇〇〇市立〇〇小学校	小学校	8	雇用者	教職		026
061	38	高校	〇〇〇市農業共済組合	農業保険業務	4	雇用者	保険契約外勤		072
062	30	高校	〇〇〇縫製KIK	婦人服製造卸	4	雇用者	営業事務	係長	055
063	29	高校	〇〇〇製作所	ガスメーター製造	6	雇用者	資材管理	倉庫班長	055
064	26	高校	〇〇〇工業	自動車修理業	3	雇用者	板金工		125
065	63	中学	〇〇〇製罐〇〇工場	缶詰用缶製造	6	雇用者	品質検査	嘱託	129
066	40	大学	〇〇〇高等学校	高校	8	雇用者	高校教員		028
067	30	中学	〇〇〇油化	化学工場	7	雇用者	運転員		217
068	47	大学	〇〇〇〇〇化学	工業薬品製造	5	重役	薬品の研究	取締役	007
069	31	中学	喫茶〇〇	喫茶店	2	自営	自営		066
070	25	高校	〇〇〇相互銀行	相互銀行	6	雇用者	受付		055
071	37	中学	〇〇〇工業	特殊機械(ダンプカー等)貸付	4	雇用者	機械の運転		233
072	23	高校	〇〇〇精工〇〇支社	自動車部品販売	7	雇用者	旋盤		121
073	64	中学		売薬行商	1	自営	売薬行商		068
074	30	大学	〇〇〇〇工業	自動車部品製造	7	雇用者	企画		055
075	39	中学	〇〇〇木材	木材市売り	3	雇用者	集金 外交		071
076	50	高校	〇〇〇靴店	靴屋	1	自営	製造 販売		192

番号	年齢 (歳)	最終 学歴	従業先			職業			職業 小分類 コード
			名称	事業内容	規模*	地位	仕事内容	役職	
077	28	大学	〇〇ペビー	貸おむつ業	4	雇用者	営業管理	営業次長	051
078	43	中学	〇〇社	製本業	1	家族	製本工		185
079	51	大学	〇〇〇病院	精神病院	4	雇用者	精神科医師		013
080	35	高校	〇〇製鋼	帯鋼パイプ製造	7	雇用者	焼鈍工		119
081	38	高校	〇〇放送〇〇支社	放送局	6	雇用者	技術専門職	主事	006
082	53	中学		煙草, 養蚕, 水稻	1	自営	煙草, 養蚕, 水稻		076
083	48	大学	国民金融公庫〇〇支店	政府金融機関	7	雇用者	管理職	支店長	051
084	35	高校	〇〇鉄筋〇〇出張所	工事の鉄筋組立	3	雇用者	現場監督		230
085	56	高校	〇〇理化器材	理学用器機販売業	2	自営	経営全般	代表取締役	064
086	27	高校	〇〇自動車工業所	自動車の板金修理	3	雇用者	修理		138
087	40	大学	〇〇食堂	食堂	1	自営	調理師	経営者	276
088	27	高校	〇〇〇〇モーター	自動車販売 修理	6	雇用者	整備		138
089	25	高校	〇〇堂	パン, 洋菓子の製造 販売	4	雇用者	製造		204
090	41	高校	〇〇〇ヤ	紳士婦人服製造販売(卸 小売)	4	雇用者	製造部管理	製造部次長	051
091	33	大学	全国漁業組合連合会	漁業組合	6	雇用者	漁業用機器の指導	技術主任	005
092	32	高校	〇〇工業所	熔接業	2	雇用者	熔接工		123
093	30	中学	〇〇精器	機械製造	4	雇用者	研磨	班長	145
094	36	高校	〇〇自動車〇〇工場	乗用車の製造	7	雇用者	生産管理	総括(係長)	055
095	57	中学	〇〇環境KK	公害防止機械の販売	2	自営	販売	社長	064
096	54	高校	〇〇市役所	市役所	8	雇用者	建築の設計監理	係長	055
097	62	中学	〇〇織物	帯地製造	4	雇用者	帯地製造		154
098	39	大学	〇〇新聞店	新聞店	2	自営	店主		064
099	46	中学	〇〇工務店	建築業	2	自営	経営主		220
100	65	高校	〇〇〇屋	袋物卸	4	重役	販売	重役	051

*従業先の規模は次の通りである。

1: なし 2: 1人以上 3: 5人以上 4: 30人以上 5: 300人以上 6: 500人以上 7: 1,000人以上 8: 官公庁

Analysis of Open Ended Questionnaires within
the SSM Occupation Data

— for Construction of an Expert System

for SSM Occupation Coding —

Kazuko Takahashi

Interpreting occupational information from open ended questionnaires is, amongst a series of occupation data, the most difficult task in construction of an expert system for SSM (Social Stratification and Social Mobility) occupation coding. Therefore, it is necessary to build a dictionary so as to enable the system to process the information.

For this purpose, I have analyzed about thousand occupation data from the SSM survey done in 1965, and have gained some beneficial results.

The gap concerning abstractness between the vocabularies of the answers to the questionnaire and those in “SSM Occupation Classification (revised edition),” which is the source for database systems and knowledge-based systems, is relatively small.

Although the former come in wider varieties, they are mostly in “IS-A” or analogical relationship with the latter.

When dealing with occupation coding, the point of observation, for example, the headings or vocabularies, varies from one upper classification to another. The same is true for lower classifications.