

職業コーディングにおける 知識ベース・システム構築の構想

高橋 和子

目 次

1. はじめに
2. 職業コーディングにおける問題
3. 知識ベース・システムの適用
 3. 1 知識ベース・システムについて
 3. 2 プロダクション・システムの採用
4. おわりに

1. はじめに

本稿は、自由回答の形態で収集される職業データのコーディングに、コンピュータを利用した知識ベース・システムを構築する構想について述べるものである。本システムを考えるに至った動機を述べるために、自由回答の処理・分析方法に関するこれまでの経緯から説明する必要がある。

社会調査で収集される情報を変数として考えると、定量的なデータと定性的なデータとに分けられるが、統計処理を行うためにはいずれもデータの構造化を行って、データ・マトリックス¹⁾を構成する必要がある(直井1983)。しかし、中には簡単にこの手続きが踏めないもの、たとえば自由回答(自由記述による回答)などが存在しており、これに関する処理・分析方法については明確化および体系化されていないの

が現状である(原1992)。

自由回答は、選択回答(提示された選択肢から回答)と比較すると明らかなように、データそのものにおもしろさがあるといわれるが、それ以上に、分析者ではなく回答者の視点から得られる情報であるという点で貴重である。従って、これまでプリテスト以外にはあまり活用されず、特に統計処理には不向きであるとされてきたが、単に回答を例示するだけでなく、情報の豊かさを生かした処理・分析の方法を模索していく価値は十分にあるはずである。この立場から研究を重ねた結果(高橋1990、1992abc)、次のような結論が得られた。

まず、イメージを尋ねる質問など自由回答でなければ得られない種類の情報が確かに存在する。これは、選択肢の作り方が不適當であるとか限りなく多くの選択肢を用意しておけば、回答はすべて選択回答で尽くせるとする意見に反論するものである。次に、統計処理を行なう回答は、妥当なカテゴリーに対応付けてコード化する必要があるが、その際、コーディングにおける信頼性と妥当性の問題が存在する。最後に、カテゴリーがあらかじめ存在しない場合には、得られた回答から適当に生成する必要があるが、これはかなり困難な作業であり、また生成できた場合でも、カテゴリーに対する信頼性と妥当

性の問題が存在する。²⁾

結局、一口に自由回答といっても、あらかじめカテゴリーが準備されているかどうかなどによりさまざまな場合があり、信頼性や妥当性の問題まで考慮すると、そのコーディングについても一律には処理できない面がある。従って、今後の方針としては、自由回答の場合分けが必要で、最初は単純な問題、すなわち「カテゴリーがあらかじめ定義されている自由回答のコーディング」についての考察から始め、そこでの知見を基により複雑な段階に進み、最終的には「収集された自由回答からカテゴリーを生成し、それに基づくコーディング」について考察する方が生産的であるとの結論に達した。もちろん、カテゴリーの個数が極端に少なかったり、その定義する領域が広範なものについては、方法論までもちだす必要がないことは明らかである³⁾。

以上を踏まえた結果、「カテゴリーがあらかじめ定義されている自由回答のコーディング」の例として、今回、「社会階層と社会移動全国調査 (SSM調査)」における職業データのコーディング (以下、職業コーディングという) を取り上げ、考察を行うこととした。詳細については次節で述べるが、SSM調査における職業データは通常のものより複雑で、自由回答と選択回答の計 6 種類から構成される。これらはさまざまな統計処理の変数として用いられるために、293種類の職業のカテゴリーにコーディングされる必要がある。しかし、ここには次のような問題がある。すなわち、カテゴリーはマニュアル (以下、職業マニュアルという)⁴⁾ で明快に定義されているものの、種類が非常に多い上 (従って、必然的に定義される領域が狭くなる)、

サンプル数も膨大であるため、コーディング作業に多大な労力を費やさざるを得ない。

これを解決するために、コンピュータと人間が協調して問題解決をはかる協調問題解決システム (Cooperative Problem Solving System) の立場に立ち、コンピュータにいわゆる「知識」をもたせて支援させようとしたものが「知識ベース・システム (Knowledge-based system) の適用である。

一般に、知識ベース・システムを汎用的なものとして構築することは非常に困難であるが、対象領域を限定すれば、比較的、実現性が高くなるとされている。今回は、職業データという非常に狭い領域を対象とするため、ある程度実用的なシステムが構築できるのではないかと判断したわけである。

以下、次節で職業コーディングについて問題点を中心に述べた後、これに知識ベース・システムをどのように適用するかについて 3 節で述べる。最後に、今後の計画を含めたまとめを 4 節で行う。

2. 職業コーディングにおける問題点

SSM調査は「Social Stratification and Social Mobility Survey」の略称で、社会学者を中心として1955年から10年毎に全国規模で実施されてきており、研究成果も多数刊行されている⁵⁾。調査規模の大きさは、例えば1985年の場合、サンプル数が約4,000であった。本人の現在の職業だけでなく、始めて就いた職業、父親などの職業もあるため、1調査票当たり平均10個の職業データが含まれる。従って、職業データの

個数は約4万程度にもなる。

職業データは、① 従業上の地位 ② 従業先の名称 ③ 従業先事業の種類 ④ 雇用者数 ⑤ 本人の仕事内容 ⑥ 役職名の6種類から構成され、このうち②、③、⑤が自由回答、①、④、⑥が選択回答である。これらのデータから得られる情報は、国勢調査で用いる職業コードを基にした293種類のコード(以下、職業分類とよぶ)をカテゴリーとするコーディングが行われて、階層移動を始めとするさまざまな研究の重要な変数として用いられるしくみとなっている。コーディングの中心となるものは本人の仕事内容であるが、従業先の名称や種類、さらに従業上の地位や役職名なども参考にされる。

コーディングの際、カテゴリーである職業分類に関してはいくつか問題点が指摘されているが、⁶⁾ 本稿での関心は前述したようにコーディング作業の改善にあるために、ここでは取り上げない。コーディング作業の問題点を、例として1985年の場合について述べる。まず、若手社会学者を中心とした20~30人のメンバーが、泊まり込みでほぼ1週間を要するほどの作業量の問題があることに加えて、回答によっては、即座にコーディングができる知識を、ごく少数の「コーディングの専門家」しか保有していない。従って、他のものは随時、彼に質問をしてはその場を解決していくという状況のため、知識の共有化がはかれず、一方、「専門家」もあちこちで同じような質問に出会うという非効率的な状況である(原1992、佐藤1992)。

これらを解決しようとして、実はすでにメンバーの一員である佐藤(1992)、都築(1992)により、データベースの考えに立脚したシステ

ムが提案されている。前者は全文検索を行える市販ソフトを用いたもので、実用的ではあるが、検索用キーワードの作り方に専門的知識が必要なこと(一般に回答で用いられる用語とカテゴリーの用語は、抽象化の度合いが異なるために一致しにくい)や、検索のアルゴリズムが明示されないなどの欠点がある。後者はSAS(Statistical Analysis System)を用いて機械的にキーワードを切り出すために、これらの欠点は克服されているが、今のところ実用に耐える段階にはない。両者は全く異なる観点から作られたものの、回答側の用語とカテゴリー側である職業分類マニュアルの記載用語との単なる一致による解決でしかないために、コーディングの妥当性や結果に対する説得性に欠けるという限界をもつ。

佐藤らのシステムのように、コンピュータを単なるデータの蓄積場所(データベース)とみなさず、より高い能力をもった知的なものとして利用することができないだろうか。すなわち職業分類マニュアルに記載されている「知識」や専門家のもっている「知識」をコンピュータに移植し、コーダーたちが行うのと同じような知的な作業をコンピュータに行わせることができれば、コーディング作業の軽減化と効率化を図ることができるのではないだろうか。このような問に対して構想したものが、次節で述べる知識ベース・システムの適用である。

3. 知識ベース・システムの適用

3.1 知識ベースシステムについて

最初に、知識ベース・システムについて簡単に説明を行っておく。

知識ベース・システムとは、知識を利用する意志決定支援システムまたは問題解決システムをいう。これをコンピュータ・システムとしてみると、プログラム構造上の基本要素として、知識ベース (knowledge base) と推論機構 (inference) をもつ。ここで、知識ベースはすべての知識を一定の形式で蓄えたものであり、推論機構は知識ベース内の知識を使って、推論機構を実行するための制御機構である。

知識ベース・システムの代表的なものは、コンサルティング・システムや設計支援システムを始めとするエキスパート・システムであるが、その他自然言語理解やパターン理解などの知的インターフェースや知識ベース型プログラムの分野もある。

知識ベース・システムでは、データベース・システムにおいてデータとプログラムを独立させているのと同様に、知識と推論機構を分離させていめため、どの知識がいつ、どこで用いられたかを適切に知ることができる。すなわち、結果に至るまでの推論を提示できる。

知識は、知識ベース・システムにおける重要な概念であるため、ここで整理を行っておこう。

一般に、知識は各分野でさまざまに定義されるが、知識ベース・システムにおいてはコンピュータで処理する関係上、形式化されたものしか対象とせず、意志決定の推論過程に直接影響を与える情報として、効果または機能として

捉えられる。従ってここでは、事実としての知識 (factual knowledge)、判断としての知識 (judgemental knowledge)、推論に関する知識 (meta-knowledge) の3つに分類されて扱われる。

事実としての知識とは、「～は～である」と表現されるもので、これには概念対象も含まれる。事実としての知識からのみ構成される知識ベースは、それが推論過程で利用されない場合にはデータベースとの違いが特にあるとはいえない。

判断としての知識とは、事実に対する判断規則や判断手続などの集まりをいう。これはさらに2種類の知識に分類できる。1つは問題領域において広く知られ、一般的になっているもの (教科書知識) で、もう1つは専門家が自然に修得したいわば経験から得たもの (ヒューリスティック) である。ヒューリスティックは専門家の内にとどまっている知識であるが、積み重ねが進んで一般的になれば教科書知識となっていく。

職業コーディングにおける知識ベースは、それぞれ教科書知識としては職業マニュアル、ヒューリスティックとしては「コーディングの専門家」から獲得した知識から構築されることになるが、これをいかにうまく行うかがポイントとなる。

推論に関する知識とは、以上に述べたような知識をどう扱うかという知識であり、他のものより概念的には上位のものであるということから、メタ知識とよばれる。具体的には、複数の知識が適用できる状況 (競合状態) においてどれを選択するかというような場合に用られるよ

うな知識である。

これらの知識はまた、別の観点からも分類できる。例えば職業コーディングにおいては、教科書知識とヒューリスティックを用いるが、これらはいずれも原理原則を述べるほど深いものではなく、表層的で浅い知識 (shallow knowledge) である。すなわち、本稿で構想する知識ベース・システムは、浅い知識を表現し利用するものである。知識にはより汎用的で、抽象的な科学的原理や法則をコンピュータ可読の形式にしたもの、すなわち深い知識 (deep knowledge) もある。対象領域を一般的にすればするほど深い知識が必要となるが、今回のように領域を非常に限定している場合には、領域依存の浅い知識だけで解決が可能であると考えられる。

知識はさらに、記述の順序と解釈の順序が一致するかどうかという点から、宣言型知識 (declarative knowledge) と手続型知識 (procedure knowledge) に分けられる。前者は事実やルールで表現されるもので、記述の順序と解釈とは一致しない。後者は、それが一致するものである。従って、問題解決のために明確なアルゴリズムが存在する場合には、後者による表現の方が処理効率がよい (通常のプログラムをみよ) が、悪構造の場合には前者による表現しかなく、一般に知識ベース・システムとして構築されるものは、宣言型知識が用いられることが多い。

3.2 プロダクション・システムの採用

ここでは、職業コーディングにどのように知識ベース・システムを適用するかについて、検討する。

まず、職業コーディングのための知識ベース・システムは、自由回答を職業分類マニュアルに記述されたカテゴリーにコーディングすることから、問題解決のタイプとしては分析型問題 (analysis oriented problems) として捉えられる。このタイプは、選択すべき仮説の集合をあらかじめ設定でき、これと選択規則の集合により作られる問題空間が限定できるために、合成型問題 (synthesis oriented problems) など他の型に比べるとモデルの作成が比較的簡単である。

次に、知識獲得や知識表現の観点から検討する。知識獲得とは問題領域の専門知識を獲得し、知識ベースを構築する仕事をいうが、職業コーディングでは前述したように、職業分類マニュアルとコーディング専門家の両者から、それぞれ教科書知識とヒューリスティックを獲得する。

職業分類マニュアルからの知識獲得における可能性については、すでに高橋 (1994) で検討したように、表層的な形式で意味が扱えるため浅い知識の獲得が容易であるとの結論を得ている。現在は、これを人手によるかコンピュータによるかの検討段階に入っている。職業分類マニュアルは最近、テキスト・ファイルとしても作成されたため、コンピュータによる知識獲得も不可能ではないが、自然言語処理技術とも関連するためにこの種の検討が必要となる。

一般に、知識ベース・システムが実用的なシステムとなるかどうかは、知識ベースの充実度

により決まるため、教科書知識だけでは不十分で、ヒューリスティックも不可欠ことが多い。この点については、専門家の協力を得られる見通しがついている。

知識は獲得された後、一定の形式で表現され、それらが集まって知識ベースを構成するが、これを知識表現という。知識の表現形式をどのように決めるかということは、非常に重要な問題である。なぜなら、この表現形式と推論形式とは密接に関連するため、表現形式の選び方でシステムの能力が大きく左右されるからである。また、知識獲得の際に、知識を提供する専門家とシステムを構築するエンジニアとの間に生じがちなボトルネックを避けるために、人間にとっても理解しやすい表現形式とする注意も必要である。

以上を考慮した結果、職業コーディングにおける知識表現はルール型が適当であると判断した。⁷⁾ ルール型は知識をルール（規則）の集合として表すもので、比較的小規模なシステムに向く。代表的なものにプロダクション・システムと（一階）述語論理に基づくシステムがある。両者を検討した結果、プロダクション・システムを採用することとした。その主な理由としては、後者が厳密な三段論法を駆使するのに対して、前者は「もし～ならば」というような人間の日常的な知識構造に比較的近い表現形式をとることから、職業分類マニュアルの記載事項や専門家のヒューリスティックを表現するのにより適していると判断したためである。ただし、プロダクション・システムには、一つの知識に対して一つの解釈しか与えられないという欠点があるために、より複雑な判断を行うときには

ルールを構造化するなどの工夫が必要となろう。

プロダクション・システムはさらに推論方法の違いにより、AND/OR木構造型と認識一行動サイクル型という全く対称的な2つのシステムに分けられる。前者はいわゆる後ろ向き推論を行うもので結論駆動型といわれ、後者は前向き推論を行うものでデータ駆動型といわれる。最近、これらは本質的には同一のものであるとの解釈が生まれ、両者を併せもつシステムも提案されているが、今回は研究の目的上、認識一行動サイクル型プロダクション・システムとして構築を行う。

認識一行動サイクル型プロダクション・システムの構成を図3.1に示す。認識一行動サイクル型プロダクション・システムにおいては、知識ベースは「1 F 条件 THEN 行動」なる形式のプロダクション・ルールの集合からなるルール・ベースと、事実の集合であるデータ・ベースから構成される。推論機構は、データ・ベース内の事実に対してルールを適用し、更新を行うためのインタプリタである。知識の種類でいえば、データ・ベースは事実としての

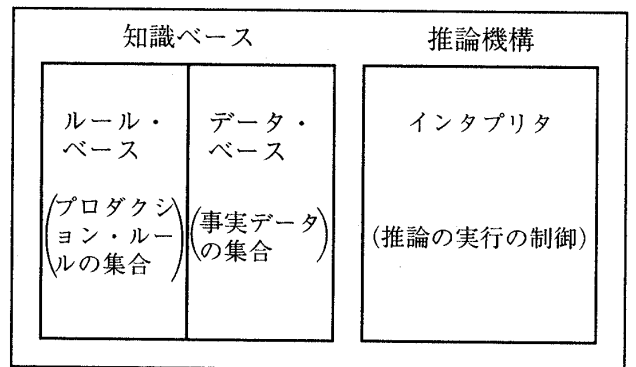


図3.1 認識一行動サイクル型プロダクション・システムの構成要素
(出典『知識工学入門』（上野1989）P74)

職業コーディングにおける知識ベース・システム構築の構想

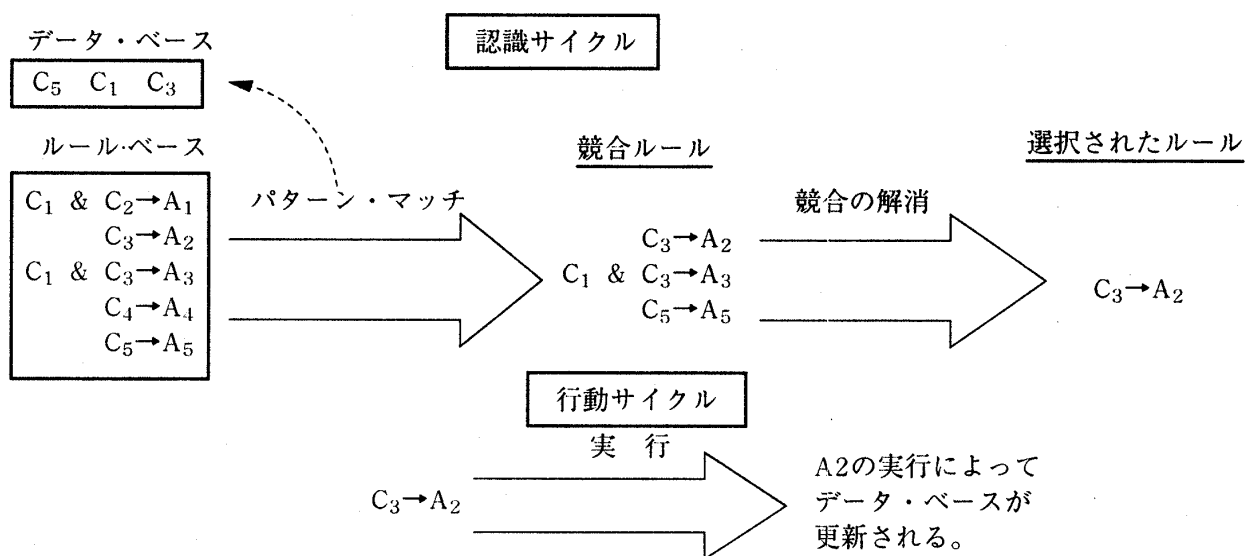


図3.2 認識-行動サイクル型プロダクション・システム推論のしくみ (出典『前掲書』P75)

知識、ルール・ベースは判断としての知識をそれぞれ管理し、インタプリタはメタ知識となる。

これを職業コーディングに適用すると、ルール・ベースは職業分類マニュアルに記載されている内容と専門家のもつヒューリスティックであり、データ・ベースは自由回答であると考えられる。すなわち、自由回答は現実世界における事実として存在し、職業分類マニュアルや専門家の知恵は、どのカテゴリーにコーディングするか判断として存在する。

認識-行動サイクル型プロダクション・システムの推論のしくみを図3.2に示す。図3.2でわかるように、プロダクション・システムにおける推論とは、認識-行動サイクルを繰り返しながら、初期状態にあるデータ・ベースを目標へもっていくプロセスであるが、データ・ベースやルール・ベースを前述のように仮定すると、これは職業コーディングにおける自由回答(初期状態)からコード化(目標状態)のプロセスであると解釈できる。

ここで、図3.2ではデータ・ベースはC₁なる記号で単純文字列として表現されているが、職業コーディングにおいては信頼性や妥当性、さらに説得性を高めるため、より強力に構造化を推め、「対象-属性-値三つ組(object-attribute-value triple)」の形をとることとする。これは、自由回答がある程度構造化されて収集されることと、高橋(1994)で検討したように、職業分類マニュアルも格文法による構造化が可能であることから実現できるものと思われる。

非常に簡単な例ではあるが、表3.1に一例を

対象“0 0 1”(自然科学系研究者)	
属性	値
場所	研究施設
対象	研究、試験、検定、分析、 鑑定、調査

表3.1 職業分類マニュアルにおける「対象-属性-値三つ組」の例

示しておく。

以上より、システムの構築のための開発言語としては、代表的なルール型の知識表現言語 (Knowledge Representation Language : KRL) の 1 つである OPS を用いるのが最適であると判断できる。ここで知識表現言語とは、当初は知識ベース・システムとして開発されたものが、より汎用化して知識表現用の言語とされたもので、プロダクション・システムの場合は EMYCIN、EXPERT、OPS などがある。今回 OPS を選択したのは、一般的にあって、最も利用しやすい環境であるパーソナル・コンピュータ上でも稼働することと、他言語を使用する可能性も否定できないので、それとのインタフェースに優れていることなどによる。

もちろん、システムの開発を LISP など従来型の言語から行うことも不可能ではないが、知識表現言語はすでに推論機構を備えていることから開発の効率性が高く、また専門知識を抽出しやすいことなどもあって、適切な知識表現言語が存在する場合にはこれを用いる方がよい。

4. おわりに

本稿では、SSM 調査の職業コーディングにおける知識ベース・システムの構築について述べた。構想内容を要約すると、知識獲得は職業分類マニュアルと専門家から行い、知識表現は認識—行動サイクル型プロダクション・システムによる。システムの開発はパーソナル・コンピュータ上で、ルール型知識表現言語 OPS を利用して行うなどである。

今後はこの構想を具体化し、実際にシステム

を構築していくことになるが、そのためには、まず知識獲得と知識表現方法をより詳細に検討する必要がある。

次に、本稿では特に触れなかったが、辞書の問題を検討する必要がある。なぜなら、前述したように自由回答と職業分類マニュアルでは、用いられる用語の抽象度が異なる場合が多いため、パターン・マッチをうまく行うためには、自由回答を一度、辞書を通してからデータベースを構築するか、またはデータベースを自由回答のまま構築しておいて、ルール・ベースとのマッチングを行う際に辞書を通すかのいずれかの方法をとらざるを得ないからである。

これらの検討事項も含め、今後の計画としては、最初は仮説が 5~10、ルールが 100~200 程度のフィージビリティ・プロトタイプを試作を行い、典型的なデータによるテストを行う。次に仮説、ルールともに増やしたプロトタイプの開発に入り、トレーニング・データその後いわゆる汚いデータによるテストを行う予定である。各段階での評価は、実用性、能力、わかりやすさの 3 項目に基く。

知識ベース・システムを調査データに適用するという発想は、ユニークではあるがそれだけに参考となる例が少ない。本稿での構想が実現できれば、対象領域を拡大したり、変更しながら積重ねを行い、将来的には自由回答の一つの処理・分析方法としての汎用性を確立していきたい。

注)

1) データ・マトリックスとは次のように行列の形にできるものをいう。ただし U_i ($i = 1, \dots, n$) は分析の単位、 V_j ($j = 1, \dots, m$) は変数、 X_{ij} は各変数に関する各単位の測定値である。

U_i	V_1 . . . V_j . . . V_m X_{i11} . . . X_{ij} . . . X_{im} . . . U_i X_{i1} . . . X_{ij} . . . X_{im} . . . U_n X_{n1} . . . X_{nj} . . . X_{nm}
-------	--

- 2) カテゴリーの妥当性については、分析者の視点による違いという説明がなされることがある。
- 3) コーディングの信頼性を高めるために、回答が複数のカテゴリーにまたがるときにはそれらを統合して、カテゴリーの個数を減らしたり、定義する領域を広くしたりすることが行われる。そのようなことが容認できる場合については、自由回答ではあってもコーディング作業は楽であり、方法論までもちだす必要はない。
- 4) 『昭和60年国勢調査に用いる職業分類』(総務庁統計局・統計センター刊行)をいう。記載例を次に示し、職業コードと職業名一覧を付録に示す。

001 自然科学系研究者
 研究所、試験場、研究室などの研究、試験、

検定、分析、鑑定、調査などの専門的、科学的な業務に従事するものをいう。

ただし、専門的・科学的知識と手段を生産に0応用し、生産における企計、管理、監督、研究など科学的な業務に従事するものは「技術者3~12」に分類される。

○物理学研究員、. . . .、○解剖学研員

×化粧品製造技術者、. . . .、×試験工

5) 例えば、1985年のSSM調査からは、現代日本の階層構造(全4巻)として、「社会階層の構造と過程」、「階層意識の動態」、「教育と社会移動」、「女性と社会階層」(いずれも東大出版)が刊行されている。

6) 例えば、職業分類の分類が、現状を反映していないとの批判もある(間淵1993)。研究の持続性を考慮するとその可能性は低い、もし将来的に分類方法が変化したとしても、知識ベースシステムにおいては知識ベースと推論機構を独立させているため、対応は容易である。

7) 知識表現には、ルール型以外にもフレーム型など5種類のものがあるが、大別するとルール型とネットワーク型の2種類に集約できる。

[参考文献]

Creedy, H. R. et al. 1992. Trading MIPS and Memory for Knowledge Engineering. *Communications of the ACM* Vol.35, No.8: 48-63.

原 純輔. 1993. 『SSM職業分類(改定版)』(文部省科学研究費補助金研究結果報告書)

原 純輔. 1992. 「定型データと非定型データ」

- 『非定型データの処理・分析法に関する基礎的研究』(文部省科学研究費補助金研究結果報告書) : 1-14
- 原 純輔・海野道郎. 1984. 『社会調査演習』 東京大学出版会
- 絹川博之他. 19. 「日本語情報検索システムにおけるキーワード自動抽出」『日立評論』 Vol.64 No.5 :
- 間淵領吾. 1993. 「職業の変化と職業小分類の問題の問題点」『第16回数理社会学会大会研究報告要旨集』
- 長尾 真. 1986. 『機械翻訳はどこまで可能か』 岩波書店
- 直井 優. 1993. 「社会調査の設計」『社会調査の基礎』 サイエンス社
- 佐藤嘉倫. 1992. 「職業コーディング支援システムの構築」『非定型データの処理・分析法に関する基礎的研究』(文部省科学研究費補助金研究結果報告書. 研究代表者 原純輔) : 199-204
- 高橋和子. 1990. 「質的データの解析に関する一考察-政治意識調査における自由回答法の分析-」『茨城大学人文学部紀要(社会科学)』 23号 :
- . 1992a. 「日本人における政治リーダーのイメージ-自由回答によるデータの処理・分析-」『非定型データの処理・分析法に関する基礎的研究』(文部省科学研究費補助金研究結果報告書. 研究代表者 原 純輔) : 137-164
- . 1992b. 「参加者からみたセミナーのイメージ-自由回答法によるデータの分析から-」『苦悩する自己啓発セミナーの研究』(文部省科学研究費補助金研究結果報告書. 研究代表者 井上芳保) :
- . 1992c. 「自由回答における構造化支援システムの開発について」『茨城大学人文学部紀要(社会科学)』 25号 :
- . 1994. 「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』 16号 : 127-140
- 田中 博・下井優一. 1987. 『エキスパートシステム構築の方法 OPS83プログラミング入門』 パーソナルメディア
- 都築一治. 1992. 「職業コーディング自動化システムの試験的構築」『非定型データの処理・分析法に関する基礎的研究』(文部省科学研究費補助金研究結果報告書. 研究代表者 原 純輔) : 205-214
- 上野春樹. 1989. 『知識工学入門』 オーム社
- 上野春樹・小山照夫. 1988. 『エキスパートシステム』 オーム社
- 吉田政幸. 1993. 『分類学からの出発』 中央公論社
- 付録** 職業分類における職業コードと職業名一覧(小分類のみを一部抜粋)
- 1 自然科学系研究者
 - 2 人文・社会科学系研究者
 - 3 鉱山技術者
 - 4 金属製練技術者
 - 5 機械技術者
 - ⋮
 - 54 管理的公務員

職業コーディングにおける知識ベース・システム構築の構想

- 55 会社役員
- ⋮
- 61 一般事務員
- 62 会計事務員
- ⋮
- 70 小売店主
- 71 卸売店主
- ⋮
- 83 農耕・養蚕作業者
- 84 畜産作業者
- ⋮
- 97 採鉱員
- 98 採炭員
- ⋮
- 105 電気・ディーゼル・蒸気機関士
- 106 電車・気動車運転士
- ⋮
- 121 制工・製鋼工
- 122 非鉄金属製練工
- ⋮
- 269 自衛官
- 270 警察官・海上保安官・鉄道公安官
- ⋮
- 274 家事手伝い（住込みの女子）
- 275 家政婦
- ⋮
- 293 分類不能の職業

ABSTRACT

Construction of a Knowledge-Based System for SSM Occupation Coding

Kazuko TAKAHASHI

SSM (Social Stractification and Social Mobility Survey) collects occupation data for individuals which contain natural language responses to questions.

It is a very hard work to classify each of them into about 300 categories by hand. For example, It required 140 person-days in 1985 SSM.

This paper proposes a knowledge-based expert system which is driven by knowledge extracted from human expert. It is a cooperating problem solving system and uses a production rule.