

序章 研究の目的 (川端 亮)

近年、パーソナル・コンピュータの発達は、目覚ましいものがある。10年ほど前から次第に浸透してきたワープロの使用とともに、この数年は、インターネットによるブラウジング、メールの交換が、急速に身近なものとなった。しかし、パソコンの使い方、生かし方は、まだ十分ではない。急激なハードの発達に比べれば、ソフトもパソコンを使う人々の意識も、まだまだ後れているといえよう。この研究は、パソコンを生かしたデータ収集、分析方法の開発のための試みである。

1. 社会調査におけるコンピュータの利用

社会調査は、質問紙を用いる統計調査法と観察やインタビューによる事例調査法とがあるが、コンピュータが利用されるのは、主に統計調査法においてである。

大量に集められた質問紙を集計する際、コンピュータはその威力を発揮する。質問紙の回答を数字のデータとし、コンピュータに入力するのは手間ではあるが、ひとたび数字としてコンピュータに蓄えられたデータは、全体の集計でも男女別などの部分の集計でもいかようにも集計が可能であり、また、そこから複雑な統計的手法による分析も可能である。10年以上昔、まだ、コンピュータがパーソナルなものとなっていなかったころ、すでに大型計算機上のSPSSやSASなどのソフトを使えば、統計的な分析を行うことはできた。しかし、そのころは大型計算機上のSPSSでは、メモリーの制約があり、たとえばしばしば用いられた林の数量化理論では、あまり多くの変数を投入することができなかった。当時、事例的な調査法に対して、統計的調査法は、サンプルは多数を扱うことができるが、限られた変数しか扱えず、多数の変数との関連を見るには、事例的調査法の方が優れているといわれたものである。しかし、現在では、それほど多くの変数を集めること自体には多少の困難が伴うにせよ、事例的な手法で人間が頭で考え得る最大限の多数の変数よりもさらに多くの変数を統計的手法として投入することは容易である。また、膨大な繰り返し計算を必要とするような分析法、たとえば最近よく使われる共分散構造分析の手法は、コンピュータの助けなしに行うことはできない。また、これらの分析もパーソナル・コンピュータが比較的安価になり出回り始めた10年ほど前においては、単純なモデルであっても1つの計算に1時間ほどの時間がかかることも珍しくはなかった。ところが現在のパソコンとソフトを用いれば、同じ計算が数秒とかからないであろう。このような環境が多くの人の利用を可能とし、そのような分析法を用いて、新たな発見を導く調査研究を生んだことは間違いない。このようにコンピュータの、特にパソコンの発達によって、もちろん弊害もあるが、統計的手法を用いた調査研究は、大きく発展したといえるであろう。

ところが、質問紙を用いた調査は、あらゆる回答形式において目覚ましい進歩を遂げた

のではない。コンピュータの恩恵を享受したのは、簡単に数値化して扱える形式の回答、すなわち選択肢法によるものである。もう一方の自由回答法は、現在も10年前とほとんど変わっていないといえる。それは、自由回答法が、調査法として選択肢法に大きく劣っているからではない。選択肢法にも指摘されている欠点が多い。選択肢による回答は、調査対象者にあらかじめ選択肢としてその回答を用意することによって、回答者の自由な回答を損なっている危険性が考えられる。また、画一的な回答を求めることから個々人の体験や心の奥深くにまで分け入ったデータを収集することができない。さらにそもそも選択肢を作成することができないような探索的な調査も多いのである。それに対して、自由回答は、回答者のまさに生の声を聞くことができるという、選択肢法にない長所がある。本来ならば選択肢法と自由回答法が併用され、調査研究に実りをもたらすべきであろう。しかし、現在の質問紙を用いる調査において、そのほとんどが選択肢法であり、自由回答法が用いられることは、きわめて珍しい。生年月日の記入のように単なる数字の記入以上の、文章としての自由回答を用いた分析は、選択肢法の分析に比べればほとんどないといってもいいくらい少ない。また、多くの社会調査の教科書においては、選択肢の作り方には、1章をさくことも珍しくないが、自由回答法についてはせいぜい数行にわたってふれてある程度で、それに1章を要することはない。その理由は明確で、自由に記述された文字をコーディングという作業において、数字に直すということが、困難であるからである。多くの場合は、いくつかのカテゴリーを作り、人手によって、多数の回答を分類していくのである。これは、もちろん多大な労力を要する。しかし、労力だけの問題であれば、何年間にわたってでも行うべきであろう。しかし、この問題は、時間によっては解決されない。多大な労力を費やしても客観的な基準によって信頼性のあるコードを与えることは、困難である。一人でこつこつとやった場合は、時間の経過とともに本人の判断基準が揺らいでくる危険があり、多数の人を集めて短期間にコーディングした場合は、一人一人の分類基準が完全に一致していないという危険性がある。そして何よりもこのコーディングが果たして信頼できるものであるのかが不明なことが、一番の問題点である。現在までのところこれらの欠点をカバーする絶対的なコーディング法がなく、自由回答を用いた信頼性のある統計分析は少ない。

本研究は、いくつかの調査法によって、多数の自由回答データを集めた。そして、これらを信頼性のあるデータにするためのコーディング方法を開発し、さらにこのコーディングによってはじめて可能となる分析を探索する試みであり、さらにそれをその他のテキストデータにも応用しようとしている。つまり、自由回答法の研究であるとともに、内容分析に関する研究である。

2. 日本における研究の現状と問題点

近年の自由回答を用いた分析で目についたものとしては、原(1987)、高橋(1992a)、豊

田・前田（1994）、村瀬他（1995）がある。原（1987）では、6カ国の中学2年生相当の生徒に「日本」あるいは「日本人」をテーマとして書かせた作文を手がかりに対日イメージを分析したものであり、高橋（1992a）では、与野党の政治リーダーに対する好感と嫌悪感などについてかかれた自由回答を分析している。豊田・前田（1994）では、高校の進路指導担当教員を対象に大学入試改善案を具体的に書いてもらった自由回答を、村瀬他（1995）では、住んでいる地域にごみ処理施設が建設されるという状況を仮定し、どのような行動をとるのかを自由に回答してもらったものを分析している。これらはいずれも分析者が得られた自由回答を読んで、カテゴリーを作り、コーディングをするという手続きをとっている。いずれのコーディングにおいてもかなり注意深く行われている様子がうかがえ、その信頼性に関しては、実質的には何ら問題がないと思われるが、それを確かめる手だては示されていない。また、これらのデータの量を見てみると、原（1987）では、406の作文で、その平均文字数は800字弱、高橋（1992a）では、1,018サンプル、豊田・前田（1994）では、1,332サンプルで、平均的な回答としてあがっている1例は、およそ800字程度で、すべてのコーディングに要した時間が、延べ180時間程度という記述があり、村瀬他（1995）では、有効回収票が1,228票となっている。いずれも量的な分析が行われているが、十分分析に耐えるだけのケース数があり、分析者がカテゴリーを作成し、コーディングをするという方法も実用的であると考えられる。

このような分析者によるコーディングという方法に対して、可能な限りの多くの部分をコンピュータにコーディングさせるという方法も見られる。もっとも実用的なものに近いものとして、SSM職業コードのコーディングがあげられる。職業のコーディングは、社会階層や職業の調査においては不可欠である。簡便にいくつかの選択肢を用意し、回答者に選ばせるという方法もよく用いられているが、多種多様な職業についている回答者が自分の職業を分析者が要求するカテゴリーに正しくあてはめて答えることは現実的には困難である。したがって、職業を中心とする調査においては、職業について回答者に答えてもらった内容を、200以上にも及ぶ職業分類表にあてはめて、コーディングを行うこともしばしば行われている。この作業は、かなり熟練を要する。熟練したものは、「大学で社会調査を教えている」という回答を見れば即座に「職業コード524の大学教員」のように答えることができるが、そこまで熟練するにはかなり多数の職業コーディングを経験することが必要であろう。ところが、身近にはそのように役立つ便利な人は、あまり存在しない。そこで、周りにたくさんあるパソコンをコーディングの手助けに使うという試みが行われてきた。そこでの基本的な考え方は、「大学」や「教える」という言葉から、「職業コード524」という答えを引き出すデータベースを作ることである。佐藤（1992）や都築（1992）は、そのようなコーディングを支援するシステムを提案している。もちろん、人手を全く要しないものではなく、人間によるコーディングをコンピュータが手助けするというものではあるが、データベースを整備していくことによって、コーディングの作業が飛躍的に軽減されるというメリットがある。

もう1つの注目すべきものは、形態素解析システム JUMAN を用いた高橋(1998)の職業のコーディング自動化システムである。これは、回答の特性をうまく利用し、格フレームによる回答の意味表現を行うという方法を採用することで、コーディング・ルールを明示化し、コーディングに一貫性を持たせるという他にはない長所を引き出したコーディングシステムである。単語間に切れ目のない日本語を形態素解析により、単語に切り分け、品詞をつける機能がある JUMAN は、評価も定まっており、これを用いたコーディングシステムは、今後さらに研究が進められていくべきものと思われるが、問題点も数多く残されている。自分の就いている仕事についてというかなり限られた自由回答を対象としても、各種の辞書ファイルを作成しなければならない点、いくつかの機能を付け加えなければならないという問題点があり、さらにこの限られた回答においてさえも処理しきれない複雑な構造の回答が見られる点ことが指摘されている。他の自由回答、さらには文章のコーディングにまで拡張するには、さらに多くの解決すべき問題が生じるであろう。

その他には、和田(1985a, b)は、KJ法の手続きを援用する方法を、高橋(1992)は、「知識ベース」を作成した後、それをもとに「推論機構」を働かせる方法を提案している。なお、コンピュータ上でのKJ法に基づくような思考支援ツールは、大学や企業の研究室で開発が盛んであり、優れたものが多いが、入手しやすいものとしては、パソコン上でKJ法が行える「ISOP-KJ法 Version 1.0 for Windows」というソフトが、株式会社アイテックから販売されている。

3. コーディングにおける問題点

手作業による自由回答のコーディングに関わる問題は、コンピュータを利用することによって解決できる点が多い。コンピュータは、読み込めるデータの形式が限られるが、その形式に従う限り、大量のデータを瞬時に読み込むことができる。そして、コーディングのような作業をコンピュータが行える命令の形式に置き換えることが必要ではあるが、その形式さえ確定してやれば、1つとして間違えることなく完全にその作業を実行することができ、また、そのスピードは、驚異的に速い。したがって、読み込むデータの形式とコーディングの手続きを整えれば、大量のデータをコーディングすることが可能である。1,000サンプル程度の自由回答というデータに限らず、メガ単位の文章データも数分といった時間で処理することが可能である。そして、さらに重要な点は、その手続きは、プログラムやファイルといった形で、明示的に示すことができる。信頼性の点でも、コンピュータの利用には絶対的なメリットがある。

このようにコンピュータをいわゆる量的調査データではなく、質的なデータに用いようとする分析方法は、「CAQDA」(Computer Aided Qualitative Data Analysis)と呼ばれ、1980年代後半から、欧米ではかなりの数の研究が見られるようになっているものである(Kelle, 1995 ; Roberts, 1997)。しかしながら、日本語のデータにこれらの研究成果を応用する前

に、少なくともつぎの2つの問題を乗り越える必要があるように思われる。

まず、第一に、コンピュータが読み込めるデータという形式で、日本語は、困難な点がある。以下欧米語の代表として、英語を例にしていくが、欧米では内容分析の対象として、新聞を用い、それらの扱う内容の継時的変化の分析が見られる。本報告書でも第2部第2章では、3つの文書データを分析した。このような活字になったものをコンピュータが読み込める形で入力するにも日本語の場合は労力を要する。英語の活字の場合は、現在普通に市販されているものを組み合わせれば、コンピュータに入力することは比較的容易である。すなわち、活字をスキャナで読みとり、OCRソフトでテキストデータにすれば、ほぼ確実に読みとることができ、さらにワープロソフトでスペル・チェックをすれば、読みとり時における間違いも容易に訂正することができる。しかし、日本語の場合、現在の新聞記事でもOCRソフトでテキストデータに変換する際の変換間違いがまだかなり見られるし、さらに少し古い新聞記事であれば、旧字体の漢字や異字体の漢字が使われていることもあり、さらに読みとりにおける変換ミスは増える。そして、どこが間違っているかを示す、英語でいうスペル・チェックのような機能も最近のワープロには校正機能と称して搭載されているが、まだそれに頼るといわけにはいかない。

実際に日本語の活字がどれくらいOCRソフトで認識されるかを第2部第2章第1節で使われる真如苑の『内外時報』を例に示してみよう。使用したソフトは、イー・アイ・ロジック(株)(旧亀島産業(株))のAutotype日英Ver.4 for Windowsというものである。このソフトで、縦書きの文章をスキャナで画像として入力し、そこから文字として認識させ、テキストデータとして保存するという作業が行った。その認識例は、以下のようなものである。

六月入日、 総本部に真如常寮さまのご臨席を戴き、 全国各依処に集う代表教徒と
回線で結んで、 第三向臨時総合道場創苑会議が開催されました。 今ばで一 っの段
階を終えた当会議は、 これまでの提案を集約した布教教化、 社会貢献などの冷部門
に別れて弁科会が論置され、 さらに具体的な検討を重ねて総合道場推進に向けていく
予定です。

(誤って認識されている文字は、太字にした)

これを見ればわかるように95%近くの文字を正しく認識しているが、これをそのままデータとして使うことはできないであろう。ただ、これくらい認識すれば、当然最初からワープロ入力するよりは格段に速い。さらにこのソフトが持つ字間や行間を細かく設定し直し、誤認識した文字を正しく学習させる機能を駆使すれば認識率は上がるであろう。しかし、それでもやはり人の目によって確認し、修正するという作業は不可欠であると思われる。この日本語に入力する際の試行錯誤については、第2部第2章第4節も参照されたい。

コンピュータによるコーディング以前の2番目の問題は、語の区切りの問題である。コーディングすべきデータの最小単位は、1文字ではなく、語、あるいは文の一部である。

文字の集まりであるこの語というものの単位が、日本語の文章の場合、難しい。

つまり、どうやって文を語に区切るのか。その方法は2つある。1つは辞書ファイルを使う方法であり、もう1つは、コンピュータのプログラムで自動的に区切ることである。第一の辞書ファイルを使う方法として、本研究で取り上げるのが、AUTOCODE プログラムであり、自動的に区切るのが、KTCoder である。それぞれについての詳細は、第1部を参照していただくとして、それぞれのプログラムの基本的な考え方について簡単に説明しておきたい。

AUTOCODE プログラムは、そもそも質問紙を用いた調査で、自由回答法によるデータをコーディングするために開発されたものである。その特徴は、コーディングの手続きをコーディングルールファイルとして出力し、それを明示することが可能な点である。自由回答データをすべてテキストファイルとして入力し、そこから取り出したい語や文章の一部をコーディングルールファイルに書き出す。そして、データ中にコーディングルールファイル上の語があれば1、なければ0という値を出力するファイルを作成し、それをを用いて統計ソフトでさらなる分析が可能となる。このコーディングにおいては、コーディングルールファイルの作成にもっとも労力を要する。そこに労力をかけることは避けられないが、Window を設け、ドラッグ・アンド・ドロップの操作によって、また、コーディング結果を逐次表示することによって、その作業を少しでも容易にするようにユーザー・インターフェイスに工夫が施されている。この AUTOCODE プログラムは、今回の研究期間に Windows 版にバージョン・アップしたが、以前の MS-DOS のバージョン、他のバージョンでの調査分析が、すでにいくつか発表されている（佐藤、1993・1995；川端、1995；川端・高木、1996；豊島、1997）。

それに対して、KTCoder は、長文のテキストデータをコーディングすることを第一の目的に開発された。テキストデータから取り出したい語や文章、あるいは取り出したくない語を指定する辞書ファイルも用いるが、語の切り出しの多くの部分は、字種切りと呼ばれる単純な方式がとられ、コンピュータが自動的に行う。もちろん、この方式は、JUMAN に代表される形態素解析を使う場合に比べて、正確さは劣る。職業コードをふる場合は、何よりも正確さが求められるので、この方式を使うのは適切ではないが、大量のデータをコーディングする場合には有効であると判断し、字種切り方式を採用した。その有効性については、第2部第2章の分析例を参照していただきたい。そして、この KTCoder は、切り出した語が、テキストデータのどこに出てくるのかを示す、位置情報とともに出力され、データベースとなる。ここからすぐに統計ソフトに読み込ませ、統計的な分析をすることも可能であるし、データベース上でいくつかの処理をすることもまた可能である。データベースとしての利用価値は大いにあると思われるが、この報告書においては、そのデータベースとしての利点は、ごく単純な部分しか示すことができなかった。

さらに AUTOCODE の分析においても分析方法としての研究は十分であったとは言い難い。分析方法に関しては、今後十分に成果を見いだせる課題である。

4. 質的データの量的分析

コーディングは、コーディングするためにするものではない。分析方法、分析結果があってこそコーディングがある。つまり、分析に適した形にコーディングするのである。その分析のイメージは、不完全な形ではあるが、第2部の分析編に断片的に現れている。それは、コーディングされたデータから統計的手法を用いて、単なる自由回答やテキストデータを読んでいるだけでは、気づかない、あるいは気づきにくい回答者の「潜在的論理」を発見しようというものである。これは、回答文を読んで、文字通りに解釈するのではなく、回答文をいったん語に分解し、その語と語の間の連関の強さを元に単語の使われる暗黙の意味構造を新たな「潜在的論理」として、いわば文法のように取り出し、文章理解の枠組みとして用いようという試みである。

本研究は、自由回答データやテキストデータを用いており、いわゆる質的なデータを用いた分析といえる。質的なデータの長所の1つとして、読んでわかるおもしろさ、生々しさというもの、見田(1979:139-140頁)のいう「追体験的な了解可能性」があげられ、そこに「おもしろさ」がある。その「おもしろさ」について、原(1992:5-8頁)は、質的データのデータそのものの「おもしろさ」という点を強調している。「おもしろさ」には、他にも処理・分析過程の「おもしろさ」、分析結果の「おもしろさ」の2つの側面があり、量的データでも質的データと同じくこの2つの側面においては、「おもしろさ」があり得るが、量的データにおいては、データそのものを見て「おもしろい」と感じることはまずないであろう。質的データの「おもしろさ」は、例示によって感じられる「微妙なニュアンス」、「生きいきとした現実性」である。原が引用した見田の新聞への身の上相談の事例をあげてみる。

新潟に住む妹一家のことでご相談いたします。妹には中学二年、小学六年と、今年新入学の三人の子がいるのに先日夫が三十六歳の若さで急死。妹の夫はある問屋の運転手でしたが、退職金もなく香典を二万円もらったただけでした。妹は今までも人夫などをして貧しい暮らしをやりくりしてきたのに、これから先どうしたらよいかと手紙で泣いてきます。いなかのこととて葬儀などやかましく、力にもならぬ親類が多勢集まって一生の別れだといって酒ばかり飲んでいるそうです。生活保護の申請はしたそうですが、テレビの月賦が二万円も残っており、子どもの心を思うとすぐ売る気にもなれず、それに新入学の子が新しいランドセルや洋服など楽しみにしていたのを思い切らせるのがつらいといいます。このような母子家庭に、子どもが働けるようになるまでお金を貸してくれる機関はないものか、とつくづく思います。離れてすむ妹のためによりおチエを貸してください。(見田、1965)

この事例では、「テレビの月賦」や「ランドセル」という語によって、貧しさを感じさせる「おもしろさ」がある。大人は少々貧しくても我慢すればよいが、子どもには、せめて人並みの晴れの姿、楽しみを与えてやりたいという、新入学の子どもを持つ人にしかあり

得ない例示がきわめて「おもしろい」のである。ここにおいては、「テレビの月賦」や「ランドセル」という語は、長文のデータの中にたった1回しか出現しなくてもそれはきわめて重要な語である。また、「テレビの月賦」を「自家用車のローン」とまとめて1つの「ローン」という語に統合することも適切ではないであろう。本研究で行おうとしている分析、そこから取り出そうとしているものは、そのような質的なデータの「おもしろさ」とは異なるものである。むしろ、そのような「おもしろさ」は、切り落とす方向にある。データの中から「ランドセル」を取り出し、それを「貧しさ」を表すものとしてコーディングすること、そこには、意味理解が不可欠であるが、それをコンピュータに求めるのは酷であろう。つまり、質的データの「おもしろさ」を生かすコーディング法、分析法を考えているのではない。「テレビの月賦」も「ランドセル」も対象となる文中に1回出現したというデータセットを作り、そこからどのような分析ができるのかを試行錯誤しているのである。その点で、元のデータは質的なものといえるが、分析は質的な分析ではなく、量的なものである。

5. 本報告書の構成

本報告書は、3部構成になっている。第1部は、プログラム編で、第1章で AUTOCODE プログラムについて、第2章で、KTCoder について、それぞれそのプログラムの特徴と操作の仕方に重点を置いて、紹介している。

第2部は、実際に集められたデータの分析であり、第1章は、質問紙を使った自由回答の分析であり、第2章は、文書データの分析である。第1章では、長野県原村で行われた調査と宗教教団真如苑の霊能者を対象に行われた調査の結果からなる。第2章は真如苑の『内外時報』という教団紙の分析、天理教の『おふでさき』の分析、読売新聞の身の上相談の分析の3つで構成されている。

第3部は、印刷されたものではなく、CD-ROM で提供される、プログラム、資料である。それぞれのコーディングプログラムと分析されたデータの基礎的な集計を掲載している。

なお、添付した CD-ROM には、第3部のみならず、第1部、第2部の内容もすべて、Microsoft Word 97 の文書ファイルとして書き込まれている。

文献

- 原純輔、1987、「子どもは日本をどう見ているか」、辻村明・古畑和孝・飽戸弘編『世界は日本をどう見ているか - 対日イメージの研究 -』、日本評論社、205-218 頁。
- 原純輔編、1992、『非定型データの処理・分析法に関する基礎的研究』平成 3 年度文部省科学研究費補助金(総合研究 A)研究成果報告書。
- 川端亮、1995、「宗教教団における予言 - 「誰でも霊能力者になれる」のはなぜか」、『ソシオロジ』、40 巻 1 号、5-17 頁。
- 川端亮・高木浩人、1996、「調査例 2 質的データの分析」、小野寺孝義・山本嘉一郎編著『データ解析ミニマムエッセンス - SPSS で学ぶ統計解析』、ナカニシヤ出版、132-148 頁。
- Kelle, U., (ed.), 1995, *Computer-Aided Qualitative Data Analysis: Theory, Methods and Practice*, Sage.
- 見田宗介、1965、『現代日本の精神構造』、弘文堂。
- _____, 1979、『現代社会の社会意識』、弘文堂。
- 村瀬洋一・阿部晃士・中野康人・海野道郎、1995、「ゴミ処理施設建設政策への仙台住民の政治参加行動 - 自由回答形式非定型データの計量分析 - 」、『日本文化研究所研究報告別巻』第 32 集、37-51 頁。
- Roberts, C. W., (ed.), 1997, *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*, Lawrence Erlbaum Associates, Publishers.
- 佐藤嘉倫、1992、「職業コーディング支援システムの構築」、原純輔編、『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(総合研究 A)研究成果報告書、199-204 頁。
- 佐藤裕、1993、「部落問題に関する表現の構造 - 人権意識調査の自由回答項目の計量分析」、『解放社会学研究』7、63-86 頁。
- _____, 1995、「自由回答項目の分析」、直井優・吉川徹編『家族と高齢化社会 - 学術用モニターシステムの開発』、大阪大学人間科学部経験社会学・社会調査法講座、59-73 頁。
- 高橋和子、1992a、「日本人における政治リーダーのイメージ - 自由回答法によるデータの処理・分析 - 」、原純輔編、『非定型データの処理・分析法に関する基礎的研究』平成 3 年度文部省科学研究費補助金(総合研究 A)研究成果報告書、137-164 頁。
- _____, 1992b、「自由回答法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 巻、103-124 頁。
- _____, 1998、「自然言語処理による SSM 職業コーディングの自動化システム」、都築一治編『1995 年 SSM 調査シリーズ 5 職業評価の構造と職業威信スコア』文部省科学研究費補助金(特別推進研究(1))研究成果報告書、195-224 頁。
- 豊田秀樹・前田忠彦、1994、「大学入試方法の改善に関する進路指導担当教員からの自由記述意見の分析 - 調査研究における自由記述データの分析方法の提案 - 」、『行動計量学』第 21 巻 1 号、75-86 頁。
- 豊島慎一郎、1997、「震災転出者に関する分析 - 「西宮市からの転出者調査」から」、関西学院大学社会学部紀要 78 号、133-143 頁。
- 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」、原純輔編『非定型データの処理・分析法に関する基礎的研究』平成 3 年度文部省科学研究費補助金(総合研究 A)研究成果報告書、205-214 頁。
- 和田修一、1985a、「文書データのコンピュータ解析に関する一考察 - 理論とその適用 - 」、数理社会学研究会編『数理社会学の現在』、333-344 頁。
- _____, 1985b、「社会学における『文章データ』の解析方法に関する研究 - 問題提起と理論的背景 - 」、『社会科学討究』、第 89 号、315-341 頁。