

自由記述文の分析に対するキーワード分類法の適用

今 尚 之

【目 次】

1. はじめに
2. メディアに記録された自由記述文の分析
 - (1) 意識調査の分析対象としてのメディア上の自由記述文
 - (2) 既存研究における自由記述文の分析
 - (3) 自由記述文の数量化の困難性
 - (4) キーワードによる自由記述文の分析
3. KW 分類法を用いた自由記述文の数量化
 - (1) 自由記述文分析への KW 分類法の適用
 - (2) KW 分類法による自由記述文の分析作業
4. インハウスエンジニア問題に対する自由記述文の分析
 - (1) インハウスエンジニアに対する問題
 - (2) 分析の対象とした自由記述文
 - (3) KW 分類法の適用によるキーワードの構造化
 - (4) クロス集計によるインハウスエンジニアに対する属性別の問題意識
 - (5) 数量化Ⅲ類によるインハウスエンジニアに対する問題意識構造
5. 本研究のまとめと今後

1. はじめに

現在、社会計画の立案、実施において数々のコンフリクト問題が発生し、複雑化する問題に対して有効な解決案や合意形成を得ることが困難となっている。この背景には、従来にもました価値観の多様化が存在していることが指摘でき、社会計画の立案、実施において各層の幅広い意見を集めることが以前にもまして積極的に行われている。その中でも、アンケート調査は最も身近な意識調査方法として様々な場面で活用されている。

ところで、アンケート調査の一般的な回答方法は選択枝回答である。このため、事前に回答内容を示すことによって回答者の自由な発想を封じこめてしまうことがある。

一方、雑誌や新聞等に見られる投稿欄にはある特定の問題に対して、様々な意見や解決策の提案がなされており、さらには、近年その普及が著しい商用ネットワークやインターネット上の電子会議室などでは、同一のテーマに対し自由な立場から意見交換が行われている。これらのメディアへの投稿等による自由記述の意見は、回答に対する限定が少ないことから、自由な発想、主張が盛りこまれており、問題に対する人々の認識や問題解決のための貴重な情報源と考えられる。しかしながら、それらの意見の印象をまとめることは行われていても、実用的な集計方法が確立していないために、定量的に分析することはほとんど行われていない。

本研究の目的は、新聞、雑誌等の投稿記事や電子会議室などでの意見交換に見られる不定形な文書をキーワードによって構造化し、さらに数量化することによって、そこに見られる潜在的な

意識構造を定量的に把握する手法を構築することである。

本研究では、官公庁内に勤務する技術者であるインハウスエンジニアに関する問題を具体的な分析対象として取上げ、専門雑誌の投稿記事の分析、数量化を行い、数量化Ⅲ類によってその問題に対する意識構造を探索したものである¹⁾。

2. メディアに記録された自由記述文の分析

(1) 意識調査の分析対象としてのメディア上の自由記述文

われわれは、地球環境に関する報道が多く取り上げられるようになれば、現在の社会の関心が地球規模の環境問題に向けられていることを認識する。そして、さらに関心を持つならば、そこに書かれている記事の内容を分析し、より高度な理解を得ようとする。この例のように、われわれは、メディアに記載された記事を読み、記事内容の印象から社会全体の関心事項の推察を行い、必要な意思決定を行うことなどを日常生活の中でごく自然に行っている。

さて、メディアによって発信される不定型な情報の内容を分析し、潜在的な意識構造やメッセージの裏側に潜む内容の分析、把握を行う研究レベルでの取り組みは、内容分析(context analysis)という名称で人文学や社会科学分野においてなされてきた。ここでは、主に新聞や雑誌、放送メディアによる情報を対象として、世論や人々、特に政治的なリーダーの意識分析、さらにはコミュニケーションの在り方などの分析が行われてきた²⁾。

しかし、内容分析は数量的な集計が困難であることが問題点となり、意識調査方法として必ずしも一般的なものではない。内容分析的な自由回答形式のアンケート調査ですら、動機付けの高い回答者から質の高い情報が得られる点が認められているものの、集計の困難さから小規模な予備調査に用いられるのがほとんどである。表1にアンケート調査、内容分析それぞれの特徴を示す。

(2) 既存研究における自由記述文の分析

従来の研究においては、自由な意見(発想)を段階的に集約する方法論として、川喜田二郎によるKJ法³⁾などの構造化技法が文化人類学の分野において発達した。しかし分析者(リーダー)の資質にその結果が左右されることが多く、安定した結果が得られない問題や結果の数量的な分析が困難であることから、分析者の印象をまとめることに終わりがちである。

一方、自由記述意見の分析ではアンケート調査の自由回答の集計に関して様々な試みがなされている。大井等⁴⁾は回答内容を単語レベルに分解し、近傍法によって解析する自由連想法による調査方法を提案しており、「住みやすさ」など意味の曖昧な言葉の概念を明確化するのに有効であることが示されている。豊田等⁵⁾は、自由記述を大きな無制限カテゴリ選択項目として扱いコード化する手法を考案した。そして、大学入試の改善に関する自由記述意見を集約することを行ない、調査、研究における自由記述データの内容的、統計的分析方法を提案している。また、中岡等は回答内容を主語、述語に要約してクラスター分析を適用する方法⁶⁾や回答文書からキーワードを自動抽出し分類、集計を試みるKW分類法を提案し、北海道の道路整備に対する意識調査に適用している。

(3) 自由記述文の数量化の困難性

自由記述された文章は、図1に示した各点から、客観的な集計や統計的な処理が難しく、また大量のデータを分析することも困難である。このため、メディア上の自由記述文は特定のテーマについて記述したものに限定したとしても、雑多な内容が一文中に混在し、より混沌とした情報

表1 アンケート調査および自由記述文の内容分析の特徴

調査法	プリコード式アンケート調査	自由回答形式アンケート調査	自由記述文の内容分析
長所	1) 具体的な調査項目に沿ったデータ収集が容易 2) OCRなどの利用により迅速かつ正確なデータの収集と電子情報化が可能 3) コンピュータによる客観的集計、統計的分析が容易である	1) 動機付けの高い回答者からは質の高い情報が得られる 2) ある程度の自由度を持つため、回答があまり限定されない 3) 問題点や解決策の探索的な調査が可能である	1) 記述者の動機付けが高いため、極めて質の高い情報が得られる 2) 問題点や解決策の探索的な調査が可能である 3) 調査として目立たず、自由な発想を引出すことができる 4) 構造化しづらい資料が扱える 5) 経費はあまりかからない
短所	1) 回答範囲が限定されるので自由な発想による、意見や提案を調査できない 2) 規模が大きくなるにつれ経費がかさむ 3) 準備に時間がかかる	1) 数量的な集計が困難である 2) 回答者が調査されていることを過剰に意識することがある 3) 規模によっては経費がかさむ 4) 準備に時間がかかる	1) 数量的な集計がより困難である 2) 分析手法が確立していない

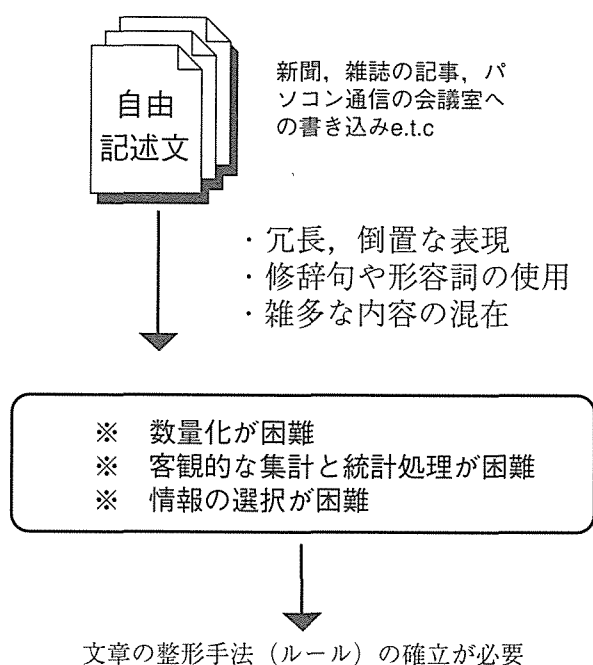


図1 自由記述文の分析の困難性

源となる。すなわち必要とする情報に対して雑音成分が多くなるので雑音成分の排除が重要となる。しかしその方法によっては必要な情報をも棄てることになりかねない。このことが、従来の自由記述文の分析を印象をまとめる程度に終わらせてきた原因でもある。このため文章の整形に関するルール作りが重要である。

(4) キーワードによる自由記述文の分析⁷⁾⁸⁾

現在、雑多で膨大にある情報の検索にはキーワードが多用されている。例えば、文献データベースなどではキーワードも同時に登録されており、検索、内容の把握に効果を発揮している。

さらに、KWIC (Key-Word-In-Context) 索引では、選択された用語 (Key-Word) をその言語環境と一緒に表示し、それぞれの用語がどのような使用をなされているかについての概念を提供することができるものである。また、テキスト中の用語の出現の観察によって、その時系

列的な変化の傾向を分析することや統計的な期待値との偏差の有意性を判断する試みなどもなされている⁹⁾¹⁰⁾。

このように、各個人の意見中の特定用語をキーワードとして整理することによって、個々の意見の共通項を探索し集約することが可能となり、キーワードに対する反応の集計によって個々の意見の特徴を定量的に把握することが可能となる。

以上より、本研究では、キーワードによって自由記述文の数量化を行ない、統計的な分析を行うこととし、キーワードの抽出に中岡等によって開発されたKW（キーワード）分類法を改良して用いることとした。

3. KW 分類法を用いた自由記述文の数量化

(1) 自由記述文分析への KW 分類法の適用

① KW（キーワード）分類法

KW 分類法はアンケート調査における自由回答方式の自由記述回答文章の分析を行うために開発された手法である。この方法は、自由回答文章からキーワード（文章中のポイントとなる語句）を抽出し、関連するキーワードの構造化（階層化）を行ない、最終的にマルチアンサー形式のプリコード項目によるアンケート調査と同様な集計を可能とするものである。また、アンケート調査を対象としているために、回答を求める段階において文章の記述方法として、短く箇条書きで記入することや記入例（〇〇を××する、△△のために□□が必要）を示すことができる。このため、回答者はある程度フォーマットに従った回答を行ない、倒置的表現や修辭句、形容詞などの使用を控えるため、キーワードの抽出作業が容易になり、分析作業が行いやすくなる。

② 自由記述文への適用

メディア上の自由記述文を調査対象とする場合、アンケート調査とは異なり、文体について指示を行うことはできない。その結果、取り扱う文章では一文中に感想、意見、提案など雑多な内容が混在し、倒置的な表現が見られる。さらに、形容詞などの使用基準が記述者によってまちまちである。このため、キーワードの抽出では慎重に文章の要約、整形を行う必要が有る。

特に、記述者が肯定的、否定的いずれの意見を述べているのかの区別、および記述者が問題に対する認識を述べているのか、解決案などの提案を述べているのか区別する必要がある。

本研究では表2のルールを定め、キーワード抽出のための文章の要約・整形作業において適用することとした。このルール作りは分析作業の結果を左右するものであるから慎重に行う必要がある。また、今回は単純化のためルールを少なくし、形容詞を考慮しなかった。将来的にはより現実的なルールの設定や形容詞の考慮などが必要である。

(2) KW 分類法による自由記述文の分析作業

KW 分類法による自由記述文の分析作業の詳細は以下である。また、図2はそのフローチャートである。

① 自由記述文原文の入力

メディア上の自由記述文の原文を省略することなくデータベースに入力する。この時、属性などもあわせて入力する。

② 自由記述文の要約・整形

自由記述文原文から、要約・整形文を作成する、

表2 本研究におけるキーワード抽出のための文章整形・要約のルール

認識型記述	A（何）は が	B（状態）である
提案型記述	A（何）を	B（行為）すべき

自由記述文の分析に対するキーワード分類法の適用

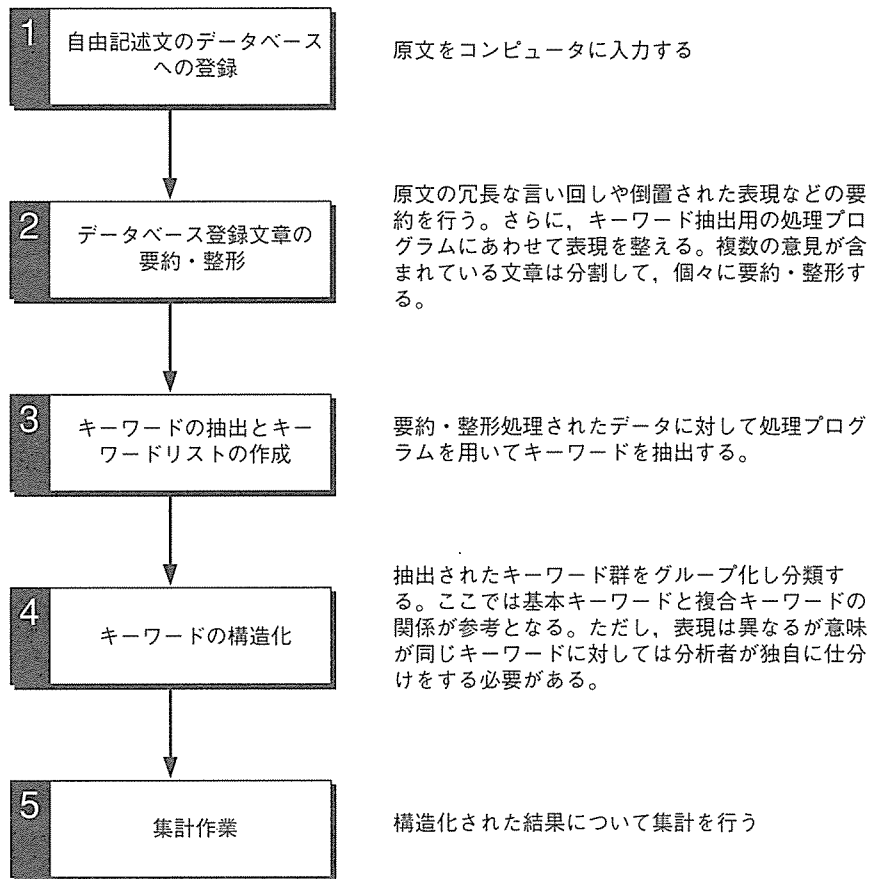


図2 KW分類法による文章処理の流れ

この作業は分析者による完全な手作業である。ただし、要約といっても倒置表現の修正、修飾語の削除を行うもので、複数内容のまとめは行わない。あたかも文章の分解を行う作業である。

③ キーワードリストの作成

要約・整形文よりキーワードを抽出し、キーワードリストデータベースを作成する作業である。本研究では表2のルールにしたがって自由記述文の要約・整形を行ったので、プログラムによって自動的に抽出することが可能となった。

④ キーワードの構造化

ここでは、③のプロセスで抽出されたキーワード群をより上位の概念でラベル付けを行ない、集約、構造化を行う。

また、下位レベルでの集約ではキーワードリストデータベースをコード順にソーティングし、プログラムによって機械的に処理をおこなうことが可能である。しかし、上位レベルでは分析者が手作業で行うこととなる。このため、本研究ではKWIC索引などを導入することによって、作業の支援を行うこととした。

⑤ 集計作業

構造化されたキーワード群を用いて、その反応に対する集計作業を行う。集計された結果は選択肢回答形式のアンケート調査結果のデータ構造と同じになるので、種々の統計的処理が可能となる。

4. インハウスエンジニア問題に対する自由記述文の分析

(1) インハウスエンジニアに対する問題

インハウスエンジニアは、一般には官公庁内において技術的な専門知識を必要とする業務に携わる職種を差すもので、例えば、地方自治体の土木課等において、公共資本の整備計画の立案や実際の設計、発注業務などに従事するものである。

現在、公共事業額の増大や民間技術の高度化、価値観の多様化、入札にかかわる制度疲労など公共事業をめぐる環境が大きく変化してきた。このため官公庁内のインハウスエンジニアに対する役割が再検討される時期に来ているといわれている。

インハウスエンジニアの今後を考察するためには、人々のインハウスエンジニアに対する意識構造を定量的に把握する必要がある。

(2) 分析の対象とした自由記述文

本研究では、インハウスエンジニア問題に関する議論が専門分野内で行われていることおよびKW分類法の適用による分析の有効性を確認することの両点から、土木に関連する専門家自身が問題についてどのような意識構造を持っているのか分析を行うこととした。そのため土木建設業界誌¹¹⁾に掲載された読者投稿を分析の対象とした。なお入力を行った投稿文は175編である。

(3) KW分類法の適用によるキーワードの構造化

本研究では分析対象を選択するために文章データベースに対して、インハウスエンジニアに関係すると思われる語句(例えば官公庁など)で検索を行なった。その結果57件の文章が抽出された。しかし、この検索はテキストデータ全てをその検索対象としたため、インハウスエンジニア問題と全く関係のない文章データも拾い上げる結果となった¹²⁾。このため、出力結果からインハウスエンジニア問題を取り上げている文章を再度選択した結果、分析対象は17件に絞り込まれた。そして、選択された文章に対しKW分類法を適用した結果、文字列照合によって抽出されたキーワードは82個であった。なお、本研究ではインハウスエンジニア問題に対する認識構造を知ることが目的としたので、表2の認識型のルールを用いて文章の要約・整形を行ない、キーワードの抽出を行った。

さらに、それらに対しては、KWIC索引や原文を参考にしながらグループ化を行った。図3は抽出されたキーワードとその上位概念を示すラベルの例である。また分類の結果キーワードは最終的に図4に示す7項目の概念に集約された。

(4) クロス集計によるインハウスエンジニアに対する属性別の問題意識

KW分類法を用いて集計された結果はプリコード形式のマルチアンサーのアンケート調査と同様な集計が可能である。本研究では図4の7項目に対して職種と年齢についてクロス集計を行った。その結果を図5、図6に示す。

全体として、業務量の多さを指摘する割合が高く、半数を占めている。

職種別に見ると、メーカーや建設業界では体制について指摘する意見が多い。また、コンサルタントでは能力を問題にする意見が60%在り、次いでモラルに対する指摘が目立つなど、インハウスエンジニア自身に起因する問題を指摘している。

さらに、公務に携わるものは業務量の多さの指摘が最も多く、次いで組織や能力を問題点として捉えおり、ほかにマニュアル化の弊害を指摘する意見も公務の方が多く見られる。

年代別では、全体的な傾向として20~40代と50~60代の間で認識している問題点が異なる。業務量の多さを指摘する意見は若い年代では半数以上を占め、特に20代では75%を占めている。

自由記述文の分析に対するキーワード分類法の適用

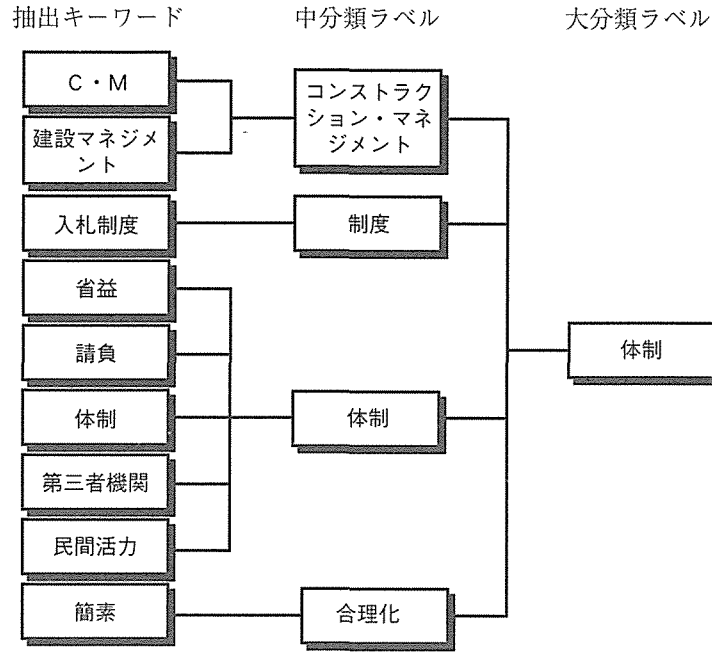


図3 キーワードと分類ラベルの例

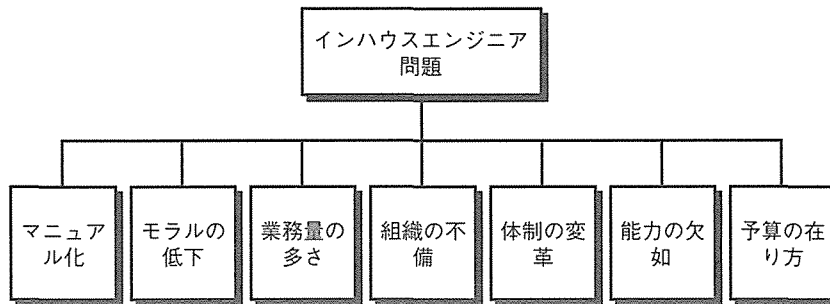


図4 インハウスエンジニア問題について集約された概念

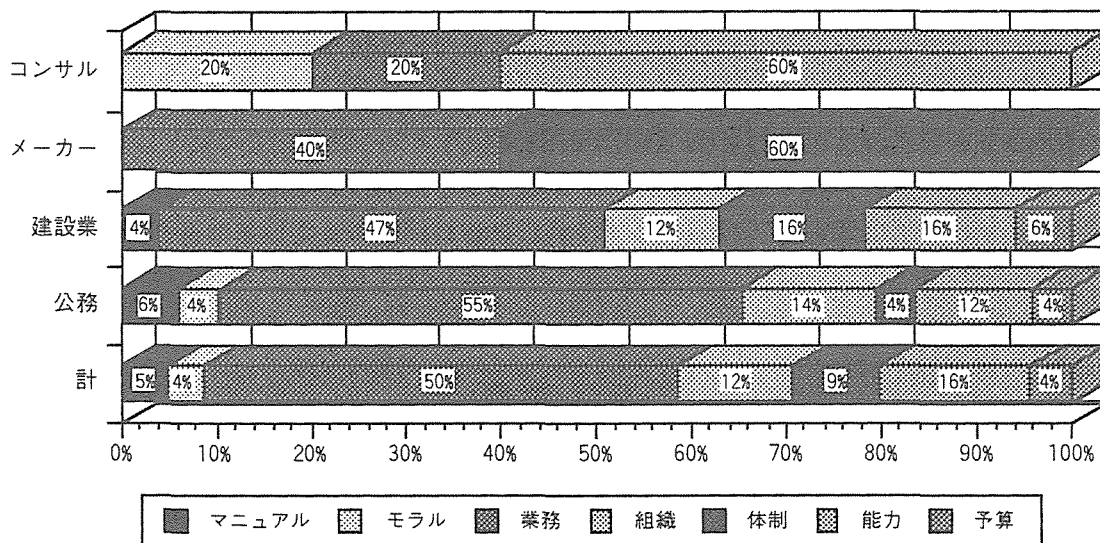


図5 インハウスエンジニアに対する業種別の意識

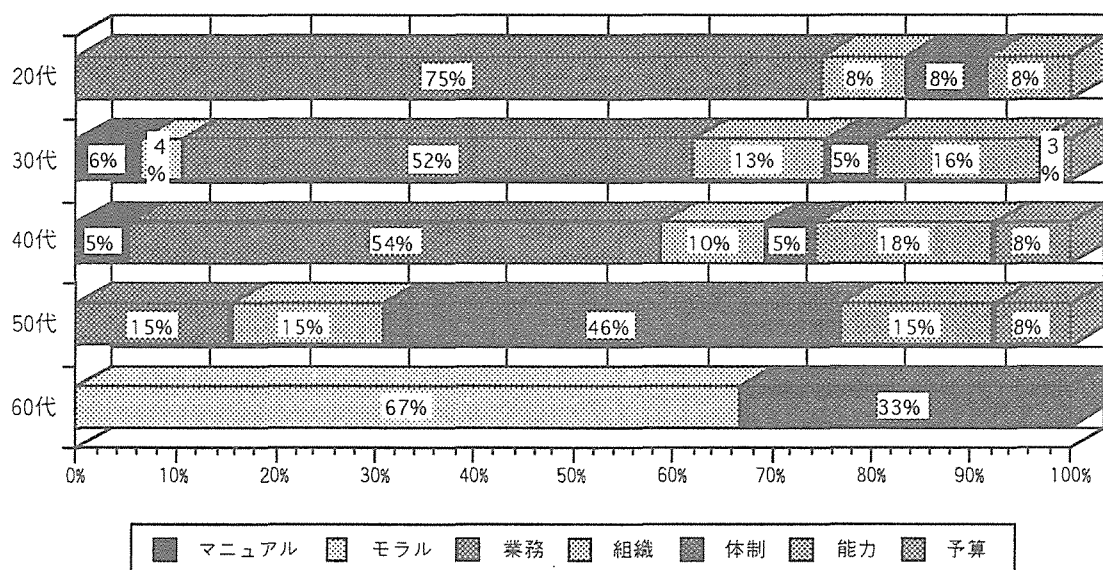


図6 インハウスエンジニアに対する年代別の意識

これは、若い世代ほど現実的な業務に追われていることと一致する傾向である。マニュアル化の弊害を指摘する意見も見られ、職種別では公務関係者が多く年代別では30～40代である。これは、この年代が若い部下を直接監督、指導する立場にあることと関係あるものと思われる。一方、50～60代では体制を指摘する意見が大きな割合を占めている。さらに、予算の在り方を問題としている年代は40～50代に見られることは、それらの年代が総合的な管理職にあることが多いことに関係するものと思われる。

以上のクロス集計結果では、おおむね妥当な傾向が現れており、本研究のデータ処理プロセスは、メディア上の自由記述文の内容分析に対して有効であると判断される。

(5) 数量化Ⅲ類によるインハウスエンジニアに対する問題意識構造

本研究では、インハウスエンジニアの現状に対し問題と認識されていることの構造を把握するために、図4の7項目をカテゴリーとして数量化Ⅲ類の適用を試みた。その結果6つの次元が得られた。また、次元の軸は個々人の内的な問題であるモラルの低下が正の最大数量となり、外的な問題といえる体制が負の最大数量であることから、インハウスエンジニア自身の問題を示す軸と考えられる。さらに、図7はI次元、Ⅲ次元両軸のカテゴリー数をプロットした結果である。図より5つのグループに分類できた。また、その結果から業務量の増大とマニュアル化そして能力の低下がインハウスエンジニア問題としてひとまとめに捉えられていることが分かった。

5. 本研究のまとめと今後

(1) 本研究のまとめ

本研究では、メディア上の自由記述文の内容分析が、意識構造の調査、分析に対して有効であることの整理から、自由記述文をデータベース化し、登録された文章の数量的分析を試みたものである。文章データの数量化にはKW分類法を適用し、データの入力から処理、分析までの作業システムを構築した。

そして、本研究の手法を実際の雑誌投稿記事に対して適用した。適用した記事は官公庁内のインハウスエンジニアに対するもので、投稿者の問題意識の把握を試みた。その結果、クロス集計

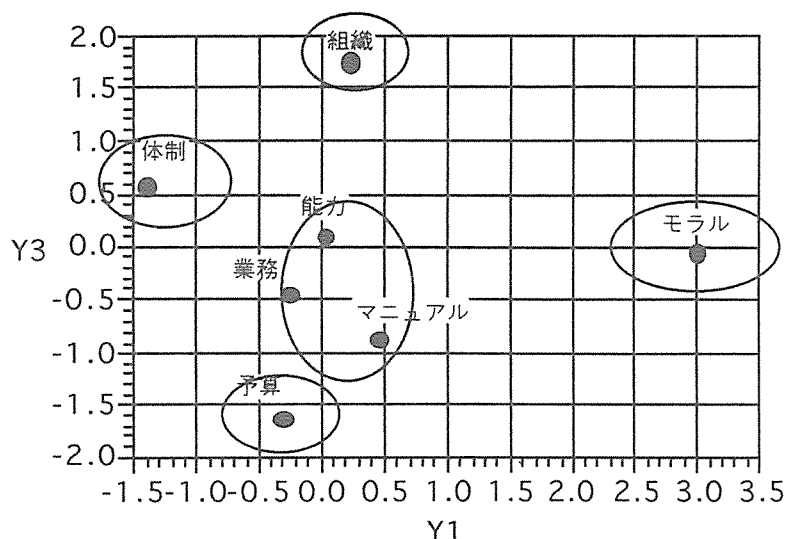


図7 数量化Ⅲ類によるカテゴリープロット図
(X軸=Y1, Y軸=Y3)

では妥当な結果が得られ本手法の有効性が確認された。また、数量化Ⅲ類を適用した。その結果意識構造は5分類され、業務量の増大、マニュアル化、能力の低下が問題として認識されていることが分かった。

(2) 本研究の今後

現在、パソコン通信などの電子会議室には、雑誌や新聞紙上の投稿記事に比べ、遙かに多彩な意見が登録されるようになってきた。この背景には、近年のインターネットブームに見られるように、コンピュータによるコミュニケーションが衆知されるようになり、さらには、パソコン通信の利用者が大手ネットワーク1社で100万人を越えるなど、大衆化、一般化が進んできたことが第一である。そして、そこに参加している人々は思いのほか幅広いものである。さらに、電子会議システムは、極端なルール違反がない限りにおいて、自分の意見の掲載がリジェクトされることがない。すなわち、発言者の意見に余計なフィルターがかからず、この点が新聞や雑誌などの投稿と大きく異なる部分である。また、時間的、空間的非同期性を持っていることに加えて、双方向性がより強いメディアであることから、意見交換による議論の発展が見られる¹³⁾。このような自由性、発展性を持ったメディア上で交換される各種の意見は、社会計画上有効な示唆に富むものと思われる。

本研究による分析手法は大量のデータを数量化できるものであり、電子会議室における意識構造分析にもまた有効なものと考えられ、今後その適用を検討したい。(1995.01.17)

参考文献

- 1) 今尚之, 中岡良司, 伊藤昌勝, 佐藤警一: KW分類法を用いた自由記述文データベースによる意識分析手法, pp.1-8, 土木情報システム論文集 Vol.4, 1995年
- 2) K. クリップENDORF, 三上俊治他訳: 『メッセージ分析の技法——「内容分析」への招待——』, pp.7-39, 勁草書房, 1989
- 3) 川喜田二郎: 『発想法』, 中央公論社, 1987

- 4) 大井紘, 宮本定明, 阿部治, 勝矢淳雄: 生活環境に関する住民認知の広がり と構造, 土木学会論文集, 第 389 号/IV-8, 1988
- 5) 豊田秀樹, 前田忠彦: 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, p. 76, 行動計量学, 第 21 巻第 1 号, 1994
- 6) 中岡良司, 森弘: アンケート調査におけるフリーアンサーの設計・解析に関する研究, 第 5 回土木計画学研究発表会講演集, 1983
- 7) 堀込静香, 中馬敏隆: 『パソコンによる書誌情報管理入門』, p. 25-35, 日外アソシエーツ, 1992
- 8) K. クリップペンドルフ, 三上俊治他訳: 『メッセージ分析の技法 —「内容分析」への招待—』, p. 188-191, 勤草書房, 1989
- 9) 中岡良司, 森弘, 五十嵐日出夫, 佐藤馨一: 文献データベースにおけるキーワード分析の適用, 第 16 回土木情報シンポジウム講演集, pp. 181-188, 土木学会, 1991
- 10) 細川寛, 高野伸栄, 佐藤馨一, 五十嵐日出夫: キーワード索引による土木用語の類型化に関する研究, 土木学会第 49 回年次学術講演概要集, 1994
- 11) 日本経済新聞社 BP: 日経コンストラクション, 1993 年 4 月~1995 年 3 月
- 12) コンピュータの導入によって, このように特に分類付のなされていないフルテキストのデータベースからも対象を検索することが可能となったことは大いに評価すべき点である。反面, しばしば対象外の情報も拾い上げることとなり, 結果として検索の効率を低下させていることも事実であり, 今後効率的な検索手法の考案, 導入が課題となろう。
- 13) 今尚之: 地域計画学的視点による電子のコミュニケーションの可能性に関する一考察, 小樽商科大学情報処理センター広報第 9 号, 1996 年