Ⅳ－１

キーワード分析による土木計画学研究の研究動向について

北見工業大学　正員　中岡良司
北見工業大学　正員　森　弘

１．はじめに

本研究は研究論文の表題からキーワードを自動的に抽出し、文献データベースに収められた研究論文群の主要な関心を明らかにしようとするものである。文献データベースには、著者らに身近であると同時に分野的にまとまりのある「土木計画学研究」の第１号から第12号までの全論文約１千件を収録した。パーソナルコンピュータを利用したデータベースの構築およびその応用に関しては、既に他の機会で発表してきたところでもあり[1)-5)]、今回のデータベースの構築では、いかに簡易なデータベースを構築するかという点に主眼をおいている。

文献データベースに収めた研究論文の題目からその論文の骨格を為すキーワードを抽出し、その頻度や相互関連を分析するという方法は、データベースの応用研究というよりも、人文・社会科学者が主に情報の伝達手段に関する深層構造を解明しようとして発達させてきたいわゆる「内容分析（ Content Analysis）」の延長線上にあるといってもよい[6)]。本研究では、約１千件の研究課題の題目をある種のメッセージと理解しその構造を把握し研究動向を解明しようとしている。

２．文献データベース

（１）データベースの構築

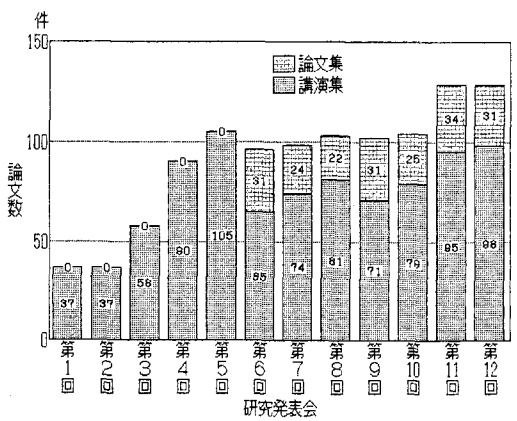
データベースの価値はその量と質に左右される。いかに質的に優れたデータベースであっても収録件数が少なくては利用する価値がない。例えば、百件程度のデータベースでは、人間の記憶力や手作業と比較して、機器操作の煩雑さを考慮すると、利用効率は決して高いものではない。本研究では、1979年１月（第１回）から1989年11月（第12回）までの土木学会土木計画学研究発表会の講演集および論文集の全論文1088件を収録したが、その選択は当該研究論文が著者らに身近であるばかりでなく、コンピュータ利用の成果が期待できる量を備えていたからでもある。

次に、質的側面では最も重要なのは入力項目である。一般的に文献データベースの項目と考えられるのは、出典、表題、著者（ときには所属も含めて）、概要、キーワードなどであり、さらに、概要を目的、方法（手法）、結果などに分けて入力する場合もある。しかしながら、文献調査といった場合、研究者の多くはまず表題を検索し、内容に関しては該当論文そのものにあたるのが通常である。本研究では、その流れに従って、データベースの項目を出典と表題に限定した。データ構造は以下の通りである。

① 掲載年：研究発表会回数で代用
② 論文種別：講演集１、論文集２
③ 掲載順：各号の論文の連番
④ 表　題：和文。英文は和訳

（２）データベースの利用

上記のデータ内容は、ＭＳ－ＤＯＳ標準ファイル形式（いわゆるテキストデータ）で作成した。本研



図－１　土木計画学研究論文発表件数の推移

A Trend of Research Themes in INFRASTRUCTURE PLANNING by Key Word Analysis
by Ryoji NAKAOKA , Hiroshi MORI

究の作業の大部分はＢＡＳＩＣプログラムを作成し処理したが、テキストデータの変換を通じて市販の多くのデータベースソフトも利用できる。本研究ではサブソフトとしてリレーショナルデータベース「桐」（(株)管理工学研究所）を利用した。市販ソフトを利用して第１回から第12回までの土木計画学研究論文数の推移を示せば図－１の通りである。

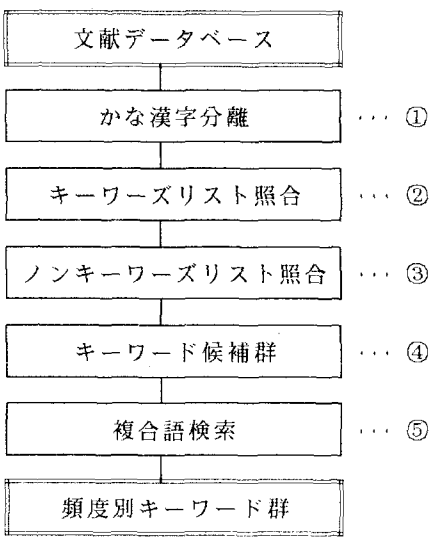
```
Free Word ? = 交通
Applicable Case =  284          Display or Printer or Continue = C
Free Word ? = 非集計
Applicable Case =   10          Display or Printer or Continue = P

02,01,025,非集計ロジットモデルによる交通手段選択の分析　－多手段同時選択の場合－,01
04,01,057,通勤交通手段の予測における集計モデルと非集計モデルの予測精度の比較,01
05,01,067,選択構造を考慮した非集計交通需要予測モデルについて,01
05,01,068,非集計モデルを用いた都市内旅客交通需要の推計方法について,01
05,01,069,非日常的交通行動への非集計モデルの適用　－チョイスベイストサンプルに対する推定問題の検討－,01
06,02,002,意識データに基づく非集計交通手段転換モデルの構築の試み,01
06,02,005,非集計行動モデルによるＯＤ交通量推計方法,01
09,01,001,非集計行動モデルを用いた総合交通需要予測システムの開発研究,01
11,01,067,非集計行動モデルに基づく都市交通需要予測体系の構築の試み,01
11,02,029,公共輸送機関の経路探索が非集計交通機関選択モデルに及ぼす影響,01
```

図－２　フリーワード検索例

さて、作成した文献データベースを利用する最大の目的は論文の検索（選択）にある。大型コンピュータを利用した汎用データベースにおいては、あらかじめ決められたキーワード（検索語）を通じて該当ケースを検索する形態が多いが、キーワードが限定されるなど利用上の制約が大きい。一方、パーソナルコンピュータ用のデータベースソフトの多くはより自由な検索を可能としているが、条件設定数に制約も多い。そこで、本研究では、指示した用語を一部でも含む論文を検索するフリーワード検索プログラムを作成した。

いま、図－２に示すように、〝交通〟という用語を含む論文を検索した結果、284件の該当があった。さらに〝非集計〟という用語を含む論文を検索した結果、10件となった。その結果は図に示す通りである。ここでは、表題の任意の場所に指示した用語が該当している様子がわかる。検索条件としては、指示した用語を含まないという排他的条件の設定も可能である。また、条件数は事実上無限である。



図－３　キーワード自動抽出フロー

表－１　キーワード・ノンキーワードリスト

[Keywords List]
1. 繰り返し　2. ふ頭

[Non Keywords List]

1. 一	2. 的	3. 法	4. 論	5. 研究
6. 考察	7. 分析	8. 考慮	9. 一考	10. 評価
11. 適用	12. 影響	13. 開発	14. 方法	15. 検討
16. 事例	17. 手法	18. 推定	19. 比較	20. 利用
21. 予測	22. 調査	23. 推計	24. 課題	25. 対象
26. 特性	27. 要因	28. 分類	29. 構成	30. 把握
31. 計測	32. 立地	33. 着目	34. 形成	35. 論評
36. 起因	37. 負担	38. 反映	39. 分割	40. 決定
41. 基礎的	42. 利用者	43. 問題点		

３．キーワード分析

（１）キーワードの自動抽出

キーワード（Key word）とは、一般に、文章などの意味を解くための鍵となる言葉を指すが、学術用語としては、その論文を最も特徴づける用語といってよいであろう。論文の表題は、その研究の目的や方法を最も簡潔に記した文章であるから、必然的にキーワードを中心に構成されている。

一般的には、言葉の意味を無視して、任意の文章

からキーワードを抽出することは不可能であるが、研究論文の表題という狭い領域に限るならば、我々は、表題におけるキーワードの切り出しを比較的簡単な基準で行うことができる。

本研究で開発したキーワード自動抽出法のフローを図－３に示す。第①段階では、文字コードを利用して表題を「かな文字列」と「漢字文字列」に分離する。基本的に漢字１文字はキーワードにならないが、例外もあるため第②段階で確実にキーワードとして抜き出したい文字列と照合する。次に、第③段階では、「研究」、「把握」などキーワードとならない文字列をノンキーワードリストとして登録し照合する。プログラム上では第②段階、第③段階は学習機能を持たせているため、処理件数に応じてより実用的なリストが作成されることになる。約百件の論文を処理した結果、得られたキーワードリスト、ノンキーワードリストは表－１に示すとおりである。以上の過程を通じて、第④段階で基本的にキーワード群が得られるが、複合的キーワード（例えば、交通量配分＝交通量＋配分）が多く、この段階では頻度は算出するには至らない。そこで、第⑤段階として、得られたキーワード群の完全相互検索によって、独立したキーワード（より最小単位のキーワード）を求め、以下の分析に利用した。

（２）主要キーワード

抽出したキーワード群のうち、上位50語をその出現頻度とともに表－２に示す。図－４は、その一部を図化したものである。図表からも明らかなように、出現頻度 100回以上のキーワード（ほぼ出現頻度は論文数と同じ意味であるが、同一論文中に２回以上使用している場合も有り得る）としては「交通」、「モデル」、「計画」、「道路」、「システム」があげられる。とりわけ、「交通」に関しては1088件中の約３割の論文で使用されていることがわかる。

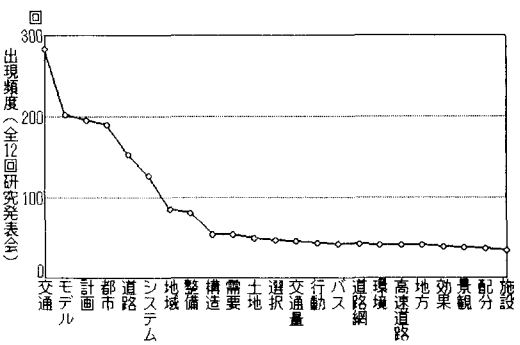
（３）キーワードの複合構造

前項で求めた高頻度のキーワードは、最小単位でのキーワードでもあるため汎用的な用語となりがちである。そこで、再び各キーワードが複合的に用いられている用語を構造的に整理したのが図－５である。ここでは、例えば、「交通」というキーワードが「交通量」となり、さらに「交通量配分」、「交通量配分計算」、「交通量配分法」と複合化していく構造が見られる。各語の末尾の数字は出現頻度である。「交通」を扱った論文が最も多いという結果があまりに一般的と考えられる場合は、その中でも「交通量」を扱った論文が多いという理解が可能である。

表－２　主要キーワードリスト（上位50語）

数字は出現頻度（回）

・交通	284	・土地利用	34	・理論	27
・モデル	202	・都市圏	33	・ﾈｯﾄﾜｰｸ	26
・計画	195	・意識	32	・情報	25
・都市	189	・住民	31	・地方都市	24
・道路	153	・高速道路	42	・交通需要	24
・システム	127	・地方	42	・解析	23
・地域	85	・効果	39	・活動	23
・整備	81	・景観	38	・港湾	23
・構造	55	・配分	37	・分布	22
・需要	54	・施設	35	・規模	21
・土地	50	・空間	34	・便益	21
・選択	48	・鉄道	34	・河川	20
・交通量	46	・街路	30	・建設	20
・行動	44	・管理	30	・市街地	20
・バス	43	・住宅	30	・人口	20
・道路網	43	・通勤	30	・ｼﾐｭﾚｰｼｮﾝ	19
・環境	42	・変動	27	（以上、上位50語）	



図－４　キーワードの出現頻度

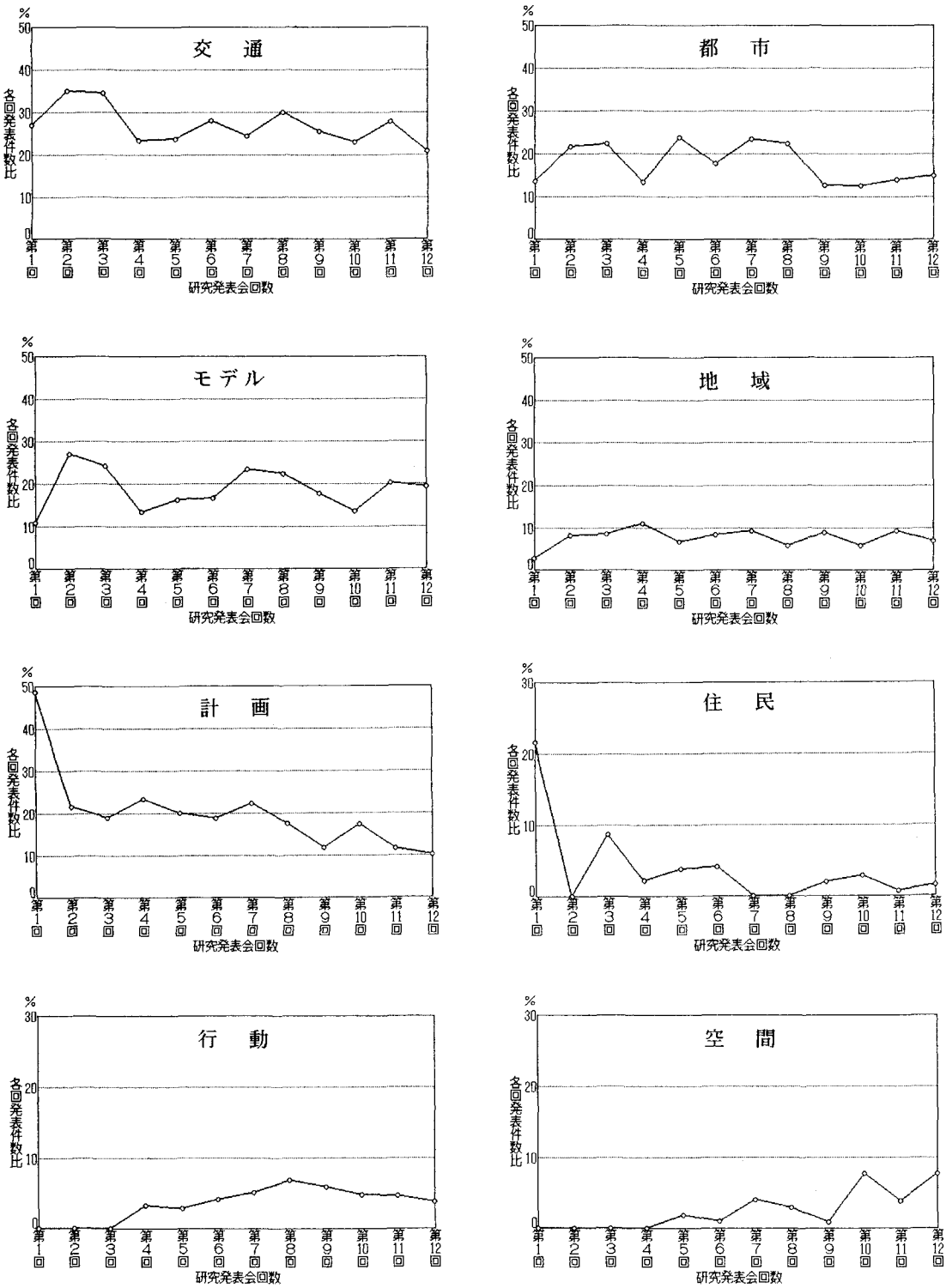
（４）キーワードの利用率

以上の分析は、第１回から第12回までの全論文を通じての結果であったが、ここでは、各キーワードの利用率の推移を見てみることにしよう。ここで、利用率とは各回の発表論文数に対する各キーワード使用論文の割合をいう。すなわち、各キーワードを使用した表題論文がその年の論文数の何割あったかを示している。

「交通」は毎年平均して３割程度を占めている。「モデル」、「都市」は増減が激しく変動している。

※　キーワード末尾の数字は出現頻度

図－5　キーワードの複合構造



図－6　キーワードの利用率の推移

「計画」、「住民」は減少傾向を示している。「行動」、「空間」はともに利用率は低いが増加傾向にある。とりわけ、「空間」という用語は第４回までの論文には一度も使われていなかったという経緯も知ることができる。

これらキーワードの利用率の推移を図－５に示した複合語のレベルまで押し進めることによって、各時代に特有の用語を顕在化することが可能と考えられるが、現時点での論文数からは特徴的な傾向を認めることはできなかった。

４．まとめ

以上、本研究では構築した文献データベースからキーワードを自動抽出し、その構造と変化を分析してきた。以下に本研究の特徴と成果をまとめる。

① 構築した文献データベースには土木計画学研究発表会の第１回論文から最近までの全論文1088件を収録した。内容は各号を示すデータと論文表題のみであるが、研究者が関係論文を検索するには十分であり、また入力作業は最小限で済む。

② 文献データベースの利用方法としてはフリーワード検索を示した。この検索法は、あらかじめ用意された用語からではなく、自由に連想した用語から論文を検索できる利点がある。

③ キーワード自動抽出法はノンキーワードリストを必要とするなど不完全なものではあるが、比較的簡単な処理によって実用的な方法論になっている。これは、例外的な場合の頻度が少ないため集計段階では表面化しないことによる。

④ 抽出したキーワード群は、「交通」や「モデル」に代表されるように、結果として汎用的な用語が上位を占めた。自動抽出により、いわゆる”あたりまえの結果”が得られたことは、反面、分析の有効性を示したことに他ならない。

⑤ キーワードの複合語構造は本研究の成果である。本研究の段階ではキーワードの意味論的処理をいっさい行っていないが、今後、関連するキーワードを体系化してゆくことによって、研究分野の構成を一層明らかにできるものと考えられる。

⑥ 主要キーワードの使用率の推移を実用的に見るには、本研究で作成した論文数では不足するようである。その制約はあっても、土木計画学研究の中にあっては減少する研究分野や新たなキーワードの出現の様子を確認することは可能である。

５．おわりに

文献検索は研究者の基本的作業であり、より簡易なデータベースを求めて試みた本研究であるが、データを単なる数字や文字列の集合としてとらえるのではなく、シンボリックな現象として分析する内容分析の立場から、表題に含まれるキーワードの分析へと本研究は発展してきた。残された課題は多いが、とりわけ、第１に、土木学会論文集や全国大会講演概要集など他の文献を対象に文献データベースを充実させていく必要がある。第２に、論文の検索方法をさらに発展させるには関連する用語グループなどの概念を導入しファジィ理論の適用が必要と考えられる。

最後に、本研究を進めるにあたり、データ入力およびプログラム開発に協力してくれた本学卒業生舟山秀樹君と学部生 中嶋道雄君に謝意を表します。

参考文献

1) 中岡良司、リレーショナルデータベースによる土木史年表の作成と応用、土木史研究・論文集、No.10、1990.6

2) 五十嵐日出夫・中岡良司、「土木工学ハンドブック（土木史年表）」、技法堂出版、1989.11

3) 中岡良司・森 弘・五十嵐日出夫・佐藤馨一、リレーショナルデータベースによる史的情報管理システムの構築と運用、土木学会第12回電算機利用に関するシンポジウム論文集、1987.10

4) 中岡良司・森 弘・佐藤馨一・五十嵐日出夫、土木史研究データベースの作成と今後の土木史研究について、第７回日本土木史研究発表会論文集、1987.6

5) 中岡良司・森 弘・五十嵐日出夫、リレーショナルデータベースによる非計量データ処理について、土木計画学研究・講演集、第８号、1986.1

6) クラウス・クリッペンドルフ、「メッセージ分析の技法－内容分析への招待－」、勁草書房、1989.8