GENERAL INQUIRER

——内容分析へのコンピュータ・アプローチ——

塩田卓和*

1 はしがき

P. J. Stone らによれば，「内容分析 (content analysis) とは，テキスト中の特殊な特徴を体系的・客観的に同定する (identify) ことによって推論するための研究技術である。」と定義されている。それは，Stone によるまでもなく，目録あるいは用語索引活動とは本来的に異なるものであるが，われわれは，内容分析の諸技術が情報検索のより高度の発展のために多くの示唆を与えてくれると考える。したがって，本稿では，内容分析のコンピュータ・システムである General Inquirer について，Stone の編著になる *The General Inquirer, A computer approach to content analysis*, Cambridge, Mass. MIT Press, 1966, 20, 651p. の概論の部分（第1～4章）を中心に紹介しようとするものである。

2 内容分析の定義

まず，内容分析の対象とは何かについて言及しよう。これは，情報の記録されたテキスト，つまり話された記録や，書かれたものをいうのであって，すべてのコミュニケーションとか，すべての人間の行動とかではない。すなわち，内容分析は，それらのすべてを分析することは不可能であり，それらがテキスト形態で記号化された部分についてだけメスを入れようとするに過ぎない。したがって多くの場合，生きた姿で人間行動を映し出すことは出来ず，静的 (static) にしか把えられない。他方，テキストは複写，共同利用，再使用の可能なこと，とくに気のすむまで分析が出来るという有利さもまた否定出来ない。

さて，静的なテキストを分析することの意味をさらに堀り下げてみよう。テキストは単語 (words) とか文 (sentences) で成り立っているが，単語は Thomas Hobbes がいみじくも述べているように，その語についてわれわれが構想する意味のほかに，社会状況の圧力を反映すると考えられる話者の性向・興味と，さらには人格の

* しおた たくわ 神戸大学経済経営研究所経営分析文献センター

特徴，表現のスタイルを含めた話者の性質を表わしている。また単語およびそのまとまりとしての文はコミニケーションと思考の日常的メディアとして機能する。したがってそれは，個人的プロセスと社会的プロセスについての基本的証拠物件であるともいえる。かくしてテキストはそれを書いた人間とそれが生産された社会状況を表現する。さらに Mead が指摘したように，言語自体が“生きもの”の如く機能し，人間の行動を規定し，状況 (situation) を発生させる。つまり，文献そのものが人間の談話と行動のかなめ (pivot) になる。したがって人間行動を理解する上で重要な要素となる各種の推論が可能となるのである。とくに社会と人間の多くの重要な変化はテキスト形式で豊富に蓄積されるので，テキストはしばしば時代の変化についての仮説をテストする最も重要な源泉となる。

実に，推論は内容分析の存在理由であり，かつ第1の機能でもある。したがって明白な (manifest) 内容を記述すること自体が目的で，かくれた (latent) 内容の解釈を第2義的にしかみない Berelson の古い定義とは対立する。内容分析はかくれた内容と明白な内容の両者を推論の過程で扱っており，単なる現象の記述にとどまるものではない。

さらに，内容分析は調査の対象となる理論と調査技術としての内容分析との統一に重点をおいており，調査されるべき理論によって，a）比較されるべきテキスト，b）カテゴリーと規則，c）測定結果から引出される推論の種類が決定される。

ここで注目すべきは，諸々の要素を含むテキストから科学的推論を導くことが，判断過程の明確化，客観化の手続，つまりカテゴリーの組立てとその同定，さらにそれを評点する諸手続によって，始めて可能になる点である。カテゴリーは調査者の理論における変数を表わしている多くの言語符号（単語，イデオム等）からなる。例えば，SELF というイメージに関心のある調査者は異なる個人が書いた文献中の SELF の言及数を同定することに興味を抱くのが当然である。そこで，I, me, mine, my, myself という一連の符号からなる一つのカテゴリーを組立てる。内容分析の基本的手続は内容特性を表わすこれらの符号がテキストの中で出現する都度同定し，それをその通りに評点することである。なお，カテゴリーをめぐる諸問題については5節にゆずる。

内容分析の手続を明細化し，体系的にデータと理論を結びつける前述の作業はコンピュータの使用によって，一大飛躍をみるに至った。つまり，コンピュータは直感的判断を明白な (explicit) 法則にかえる過程を援助して内容分析の範囲を拡大し信頼性を保証し coding の繁雑さを減少させる一方，そのスピードを保証したのである。

3　内容分析の歴史

内容分析に関する研究は歴史学，文学，哲学，社会学，言語学，心理学，精神病理学の分野でもなされてきたが，政治学・ジャーナリズムの分野での研究が主流になっていたと考えられる。歴史的には，古くから数多くの体系的内容分析の研究があるが，ここでは理論的研究が著しく進展した1940年代以後について紹介したい。

1940年代の代表的著作である *Language of Politics* (1949)を著わした Lasswell とその協力者は政治学理論の枠の中でコミニケーションを眺め，政治的神話に焦点を合わせ，政治的シンボル（例えば，自由，権利，民主主義）の出現に結びついている種々の要因の研究を政治学の重要な課題に据えた。方法論では手続の妥当性をその生産性 (productivity) に求め，測定結果の結びつきを重視し，体系的標本抽出の問題をも論じている。また応用面では体系的な厳しさとデータの注意深い分析とを結合させている。例えば，1918～1943年のソヴェトのメーデーのスローガンの分析では手続が定義され，シンボルの頻度のグラフを全体的傾向に関する注釈で示し，さらに年度ごとに政治状況と頻度の変化の関係を分析している。

第2次大戦中コミニケーション連邦委員会の外国放送情報活動分析部をはじめ，多くの機関が大量のプロパガンダを傍受し，内容分析の技術を駆使したが，日々の忙しさに追われて必ずしも体系的手続でもって行なったとはいえないようである。戦後も Lasswell, Pool, Lerner らによってスタンホード大学のフーヴァー研究所で政治的シンボルの研究が続けられた。彼らはシンボルの比較研究を通して，内容分析がどのような場合に有用であり，それがどのような洞察を与えるか，さらにはそのコスト，望ましい手続に論及した。また統計的手続の欠点，内容分析の妥当性，カテゴリーの一貫性，少数の高頻度語のカテゴリー支配，翻訳，結合体形式のカテゴリー（コンピュータのない当時では処理が困難であった），コーダーの信頼度，単位 (unit) の選択，名詞と動詞の扱い方の相違についても論じ，臨床的分析が統計的分析の前に行なわれるべきこと，とりわけコンピュータでさえ軽視出来ない予備テストの重要性を強調している。スタンホード・シンボル研究が内容分析の重要な洞察と多くの基準をつくる上に残した足跡は大きいものであったが，彼等の大規模な研究調査は継続しなかった。大規模な内容分析には手作業による限り，ある一定の限界を感じたがためで，すでに彼らは方法の将来のフィーザビリティはコンピュータに大いに依存すると考えていた。

1950年代では，言語学，心理学に関する社会科学研究会議の動き，とくに1955年の会議を無視することは出来ない。そこでは，言語の (verbal) 資料から資料以前の状態を頻度計算により推論出来るかということの問題と，単純なシンボルの頻度計

算に代わるシンボル間の内部的偶然出現 (contingencies) の計算に関する問題に議論が集中した。とりわけ後者については，何を偶然出現とみなすかを決定するために選択される規則で研究結果が左右されるだけに，重要な問題である。

会議ではこの他測定単語の問題，カテゴリー，偶然発生索引，分析プログラム，手続の標準化等が論議された。

推論の源泉として，頻度計算が意味のある尺度として使えるかは重要な問題であるが，Baldwin は Letter from Jenny の研究でその自信を深め，頻度計算はトピックに対する態度 (attitude) の強度 (intensity) とか関心の量を測定するものだと述べている。たしかに単純な頻度計算の能力を過小評価すべきではないが，改善すべき多くの問題点を残していることも事実である。その改善策には三つあり，一つにはカテゴリーの尺度 (dimension) に重みづけの目盛 (weighting scale) を適用すること，一つには等間隔目盛による単純頻度計算ではなく，目盛変更を可能ならしめること，つまりある一定のしきい値を設けて，それぞれに異なった基準に基づいて頻度を計算することである。一つには文献内容の分布状況のむらを，文献分割による計算で解決しようとする方法である。

ともあれ信頼性のある言葉のデータから推論を導くという微視的研究の歴史は測定技法の改善，データの特性を客観的・体系的に同定する分析手段の開発の歴史といってよい。General Inquirer はかかる過去の歴史から多くを学び，自らの研究技術体系をうちたてることに成功したのである。

4 General Inquirer

General Inquirer はワンセントのコンピュータ・プログラムである。その機能は a）調査者が明細化したカテゴリーに属する単語と句の実例をテキスト内部で体系的に同定する，b）これらのカテゴリーの出現と特殊な共出現をカウントする，c）一覧表を印刷し，グラフにする，d）統計的テストを行なう，e）特定のカテゴリー又はカテゴリーの結合体の実例を含むか否かによって，その文を分類し編制し直すことである。

1966年現在，General Inquirer プログラムは大型コンピュータ（IBM 7090-7094型）と小型コンピュータ（IBM 1401型），ならびに特殊な時分割コンピュータに使用出来る。

<前処理>

General Inquirer プログラムで内容分析を行なう場合，まず前処理として，分析すべきテキストと内容分析辞典と共出現摘要規則 (cooccurrence summary rules) をカードに穿孔する。ほとんどの場合直接全部のテキストを穿孔するが，テキスト・

ディスクリプタを追加する方法もとられる。テキストのコード化の適正を計る方法として，代名詞と前出の固有の名称を結びつけるための便法や，構文の位置を示すためにテキストに構文コードを付加する方法がある。前者では，Where is John? He went to Florida. という文が WHERE IS JOHN (COMPANY TREASURER) ? HE (COMPANY TREASURER) WENT TO FLORIDA (WARM VACATION PLACE) のように定義する。後者の例では，John said to Mary that Boston needs rain を JOHN/8 SAID/9 TO MARY/0 THAT BOSTON/1 NEEDS/3 RAIN/5 と定義する。ここで /1は主題を，/3は動詞を，/5は目的語を，/8, /9, /0はそれぞれ従属節を導く主節の主語，動詞，目的語を表わす構文コードである。もちろんこれだけにとどまらず，その他各種の方法が用いられるが，特殊研究を除いてあまり利用されていないようである。

次に内容分析辞典のコード化であるが，普通，辞典の各項目 (entry) が最初にきて，続いて＝印をつけ，その項目の属するカテゴリーが数字コードの形でつけられ，カテゴリーの注釈がそれに続く（第1図参照）。

第1図　辞典例

JUDGE＝21, 64　JOB　LEGAL
KEEPER＝21　JOB
LAWYER＝21, 64　JOB　LEGAL
MAGICIAN＝21, 71, 34　JOB　MYSTIC　PERFORMER
MANAGER＝21, 62　JOB　ECONOMIC

辞典の項目は必ずしも単語に限らず，イデオムとか単語のつながり (word string) を含む。

イデオムは，そのイデオムの最後の語の辞典項目の重要部分として組み込まれている指令によって，テストされる。これには二種類の指令が使用される。このうちの一つは特別の前出語を探索するようコンピュータに命令する。例えば Bats in his belfry のイデオムを処理するには，辞典項目を BELFRY＝(W, 3, BAT, 41) と定める。これはコンピュータが belfry の三つ前に bat を発見すると，カテゴリー41をあてることを意味する。さらに進んで OFF＝48 (W, 1, KNOCK, 52), (W, 2, PULL, 42, 27), (W, 2, KNOCK IT, 83), (W, 3, PAID, 64), のような定義も可能である。この場合。もしすべての括孤内の条件が充たされなければ，48の tag が適用される。

今一つの指令は tag は共出現パターンのテストのために使用される。例えば PLAY＝68 (T, 7, 48, 34) という定義は，PLAY には通常 tag 68 が与えられるが，tag 68 ではなく，tag 48 が文中で7つ前の語にすでに割り当てられていたとすれ

ば，tag 34 が代わりに使用されるという意味である，この routine はあいまいな意味の単語を，文脈中のほかの若干の単語を補足的に調べることによって，正しく同定するのに役立つ。

共出現摘要規則は tag づけがなされた後，共出現パターンを調べて，みつかるとさらに tag の追加を指令する命令系列である。例えば tag 結合77と63が出現するか，tag 結合77と84が出現したときは，tag 90をリストに加えよの場合 IF OCCUR (77) AND (OCCUR (63) OR OCCUR (84)) $ PUT (90) $ と定義する。

＜処理１＞

まず穿孔された内容分析辞典と共出現摘要規則をインプットする。なお1401型を使用する場合はディスクに蓄積される。さらにコンピュータはテキストを読取り，辞典中の項目がテキストにあるかどうか走査し，あればそれに該当するカテゴリー番号を tag として与える。この走査は一回に限らず，例えばテキスト内で knocked としてあらわれ，辞典に KNOCK とある場合，第１回目の走査では見逃すことになる。第１回目の走査のときには，その前に接尾語ルーチンですでにテキスト内の knocked は knock に代わっているので，knock のマッチングがなされ，必要な tag が与えられる。このような方法をとると，knocks, knocked, knocking のそれぞれを辞典項目に加える必要がないから辞典項目数の節減に役立つ。もちろん swimming の場合も，第１回目の走査のあと ing, m を除いてから，語幹だけのマッチングが行われる。tag づけの体系的操作のあと，tag のついた文は磁気テープに記録される (out put tape)。tag のつかない語は left over list として印刷される。out put tape は1401型により，tag が期待通りに割り当てられているかを調べるために，対照表の形でテキストと tag をリストするのにも用いられる。この場合，カテゴリー番号を自然語の tag に変換する手続をとる。

＜処理２＞

IBM 1401–1460 tag 得点手続は一文献に tag が何回割り当てられたかについての統計を提供する。その統計は表として印刷されるか，テープ上に書き込まれる。その際，一文献中でそれぞれの tag が割り当てられた単語数または文数 (raw score) と，それを一文献中に含まれる単語，または文の総数で割った数 (index score) とが用意される。文献の長さが異なる場合，raw score と index score の両者は文献の比較をするのにしばしば必要になる。例えば文献Ａが文献Ｂの２倍の長さだとして，ＡもＢも共にトピックＸを扱っているが，Ａは異なる tag が割り当てられた若クをさらに扱っているとしよう。トピックＸに関する raw tag count は二文献で同干のトピッじであっても，文献の長さに比例して頻度が調整されるならば，Ａはこれらの tag については低い index を持つことになる。このように index score を比

較することによって，言及頻度は同じであるが，他のトピックを含んでいるので，飽和 (saturation) の程度は文献Aにおいて低いということが明らかになる。

tag 得点プログラムは二つの異なる基本的カウント手続をとるが，分析者が第1に特別の言及とイメージに関心のある場合， word count が使用される。

分析者が文献に含まれる命題を研究する場合, sentence count がより有効である。また文献に含まれる文の長短により sentence-count index score の頻度評価に狂いが生じないよう，自動的に平均的な文の長さが用意される。

tag 得点表の印刷形式については，しばしば名詞，動詞等の構文カウントが同一カテゴリー内で示される。例えば，動詞としての DISTRESS のカウントが名詞としてのそれと同一行に示される。

out put tape を tag 得点プログラムによって変換することにより， tagtally out put tape が出来るが，それはさらに tag score を文献ごとに棒グラフにする Graphing program とか，標準偏差，算術平均，分布範囲を調べる統計プログラム等若干の付属的手続に供することが出来る。

<検索1>

検索質問は特別な質問式で行なわれる。tag は二桁の数字コードで表わされ，また識別コードによる検索，あるいは自然語検索（部分マッチング検索を含む）が可能であり， NOT の使用が許されている。例えば，+32−46−05/Z−(PEACE)=p4の場合，tag 32 を含んで，tag 46 および識別コードの5欄がZを含まず，かつ peace という語をも含んでいない文を探し，カウント・印刷し，テープユニット No. 4 に書き込めという意味である。

偶然出現戦略では，+41+(MEDICARE)=P, +76+(MEDICARE)=P の場合，tag 41 と MEDICARE, または tag 76 と MEDICARE が共出現する文をカウントし，印刷せよということを表わす。

ある tag が高頻度を示す場合，さらに各種の検索が可能である。そこで質問戦略を説明しよう。しばしば調査者はある tag が非常に高いカウントを示すのを見て，調査をより深めるために，可能な共出現を探求しようとする。例えばエジプトの女子学生の将来に関する自叙伝（以下E文とする）とラドクリフのそれ（以下R文とする）を比較する場合， MALE-ROLE の言及が二つのグループで高い頻度を示す。つまり，E文では， 14.8%の文数，R文では 12.1%の文数を示す。調査者は tag の共出現をさらに探求する。MALE-ROLE を含むが， FEMALE-ROLE を含まない文を+04−05=2 (tag 04 は MALE-ROLE, tag 05 は FEMALE-ROLE) の検索質問で，別のテープに書き込む。

そこでMALE-ROLE を含むE文では，R文よりJOB-ROLE, SIGN-STRONG,

SIGN-ACCEPT, HIGHER-STATUS, RELIGION の言及が多く，SIGN-WEAK, PEER-STATUS と WORK の言及が少ないことがわかる。

しかし，共出現の相違の評価には，共出現評点が予期したものと実際に異なるかどうかを調べるために，もとの tag 得点にたちかえる必要がある。例えば MALE-ROLE を含む文の subset では tag SIGN-WEAK はR文では7.4%，E文では2.4%の出現度を示す。恐らく tag SIGN-WEAK はデータの中で，これと同じように機能しているであろう。事実，SIGN-WEAK はR文全体の6.9%，E文全体の5.6%出現している。ここで SIGN-WEAK は二種類の文献で若干出現に相違があることを知るが，MALE-ROLE との共出現を考慮することによって，この相違は非常に大きくなることを知る。主な変化はE文のカウントの低下である。

このように，検索オペレーション tag と得点オペレーションの結合は，General Inquirer のオペレーションの重要な戦略をなしている。

前述したように，MALE-ROLE を含むE文はR文より tag WORK の言及に少いカウントを示しているが，その内訳はE文の8.2%，R文の 25.0%である。また全体のデーターではE文の 14.4%，R文の 18.8%である。それでは MALE-ROLE と WORK に対する関心の文脈はどうか。ラドクリフのグループについては，印刷形式の検索命令＋04＋51＝P (04 は MALE-ROLE，51 は WORK) により，夫との共働きに対するかなりの関心と夫の究極的成功に対する関心を示していることがわかる。つまり I will help my husband in his work. My husband continued to sell his works. 等である。その反対にエジプトの女子学生は自分の兄弟と未来の息子の労働に一層の関心を示し，個人的あるいは家庭的利害よりも，国家に対する貢献を強調している。つまり，After my sons graduate, they will become men who are useful to their country through great service and profitable projects. 等の実例が印刷される。

このような印刷形式での検索はしばしば，さらに検索質問を行なう上にヒントになるもので，ある検索結果をみることにより条件の強化と緩和の決定を行なうことが出来る。

以上のような調査者の戦略にあった質問式により質問に適合した文を得点と一緒に磁気テープ上に記録したり，印刷したりして，データの特性の探究とテストを行うのである。これらの検索には主として1401—1460型が使用される。

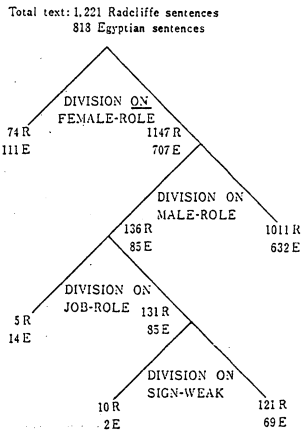
＜検索２＞

General Inquirer 時分割プログラムは二つのテキストの比較を容易ならしめるために設計されている。時分割手続はもっぱら文内部の共出現パターンに注目し，トウリー構造を用い，ある tag が存在するか否かにしたがって文を編制しなおす。

調査者はまず，内容分析プログラムを呼び出し，比較したいテキストを明示する。さらにトウリー構造戦略の記述と結びついた指令をくだす。ここでトゥリー構造戦略を簡単に説明することにする。

前述のR文献の1221文数，E文献の818文数について調べる（第2図参照)。まずFEMALE-ROLEがあるかどうかによってデータを分割する。tag FEMALE-ROLEを含むものは左，含まないものは右に示す。つまり74のR文，111のE文にtag FEMALE-ROLE が割り当てられ，1147のR文，707のE文にはその tag が与えられていない。そこで FEMALE-ROLE を含む文を分割するか，含まない文を分割するかの選択をコンピュータはせまる（これをトゥリーの分岐点での選択と呼ぶことにする)。そこで FEMALE-ROLE を含む文の分割を欲しないときは QUITをタイプする。これによってコンピュータは FEMALE-ROLE を含まない文に注

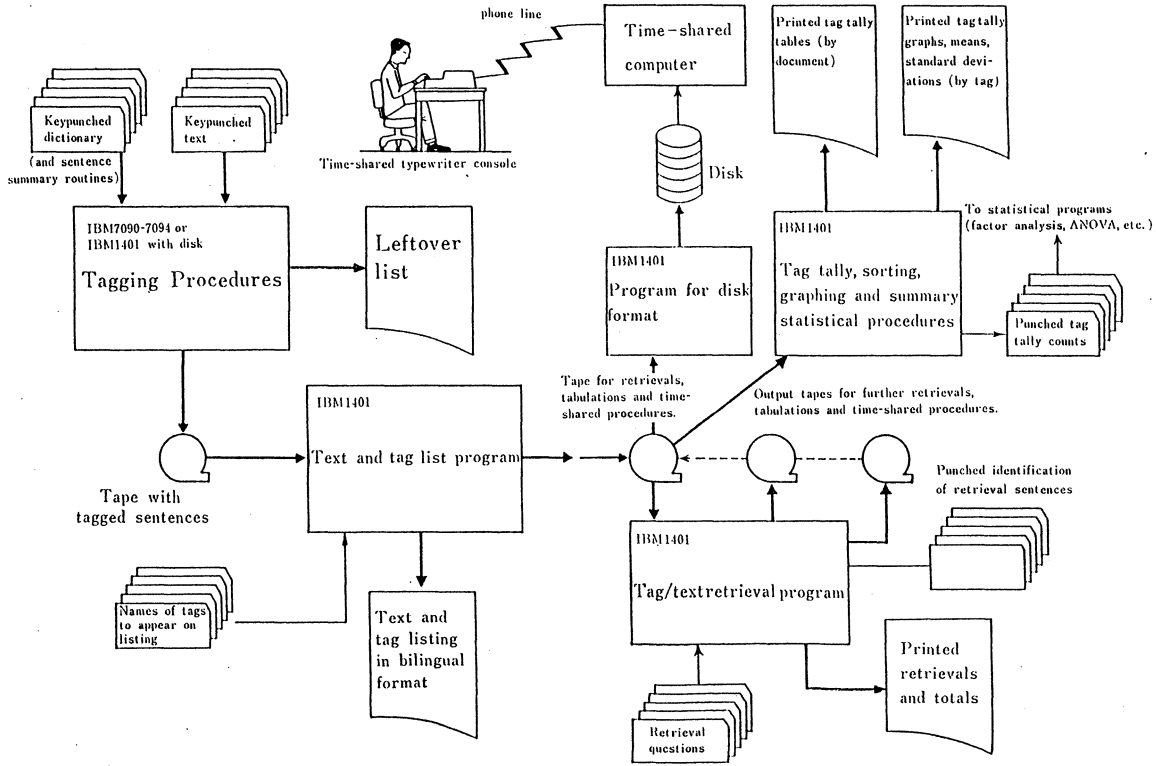
第2図　トゥリー構造例



意を向ける。MALE-ROLE に関心のある場合，MALE-ROLE を含む文を，さらにJOB-ROLE, SIGN-WEAK を含むかどうかによって分割する。調査者はこれ以上の分割は無益だと考えるところまで分割し続けることが出来る。この場合，トゥリー分岐点での選択の基準は，a）両文献とも高い頻度を示すのはどちらの方向か，b）理論的興味に適格な方向はどちらか，c）両文献の比率の高い方はどちらかである。

調査者はこのような検索戦略の過程で，印刷したいときは ORDER，主語としての tag 34 を含む文をカウントしたいときは USE 34S というような簡単なタイピングの指令でこと足りる。

第3図　General Inquirer・ブロック・ダイヤグラム



以上のようなトゥリー構造の分肢点で，調査者が判断し命令する方法に対して，一切をコンピュータにまかせる自動選択方式がある。例えば －05＋04－10－76＝P2 と最初に定義すうならば，上の例と同じ作業を行なってくれる。しかし現段階では柔軟性にとぼしく，あまり用いられない。

以上概要を述べる中で明らかなように，調査者は分析のあらゆる段階でデータに働きかけ，一つの段階の成果を次の段階での実行計画に利用出来るとともに，何回でももとのデータにたちかえって，しかも必要とあらば新しい辞典で分析を始めることが出来る。General Inquirer の特徴はまさにこれらを保証するデータへの接近の容易さ，手続の比較可能性，ならびに使用の容易さにある。最後に General Inquirer のブロック・ダイヤグラムを第3図で示しておく。

5　カテゴリーと内容分析辞典

内容分析の成否はカテゴリーの設定にかかっており，それはテキスト内の特性を理論に結びつける重要なステップである。カテゴリーは調査者の理論の変数を示す言語符号であるから，普通単純なカテゴリーが用いられることは少なく，調査者はカテゴリーの関係に興味を抱くので，辞典と呼ばれるカテゴリーの集まりが使用される。

ここでいう辞典についてその特徴を述べると，第1に標準辞典よりもシソーラスに似て，意味論的分類がなされている。第2にシソーラス，標準辞典に比較して，構文の位置で意味が変り，同形で意味が異なる場合の処置が不十分である。Harvard Ⅲ辞典はそれを克服しつつあるが，なお今後の課題でもある。第3に，分類の原理 (reason) は科学者に対するその有用性，すなわちその理論と仮説に適するかどうか，分析の実際的使用のための理論と仮説に適するかどうかによってきまる。第4に，意味論的な単位としての日常的な denotative なカテゴリーを使用している点ではシソーラスと似ている。

次に辞典の種類であるが，General Inquirer では17の辞典がすでに用意されている。それは一般辞典と特殊辞典と中間的な辞典に分けられる。一般辞典はカテゴリー数が多く，一般的理論の操作化の発展，テストを目的とするもので，Harvard Ⅲ辞典が代表的である。特殊辞典はカテゴリーの数が少ない辞典で，欲求達成辞典が代表的である。中間的な辞典は項目数が非常に多いが，カテゴリーは数個というような辞典で，その代表的なものはスタンホード政治辞典である。

辞典の編成方法の要点は調査者の理論がカテゴリー数を決定し，カテゴリーの明細化のレベルをきめ，抽象のレベルをきめるということである。最近よく行なわれているのは，分析対象のテキスト中から標本をとり出し項目語をきめたり，カテゴ

リーの抽象レベルをきめるという方法である。一般辞典の編成には主として，カテゴリーに subsection をもうけ，コンピュータに得点をカウントさせて，それによってカテゴリー・レベルを決定する方法がとられる傾向にある。辞典はこのようにして，理論の側と具体的データの側の要求から生まれた産物といえよう。

カテゴリー編成については，まずそれを困難にしているものに高頻度語の問題がある。ある特定の高頻度語がしばしば，一つのカテゴリー内で圧倒的なカウントを占めるので，他の項目語のカウントを無意味なものにし，ひいてはカテゴリー評価を誤らせる可能性がある。このような高頻度語に限って多くの異なる意味をもっている。この手だてはあいまいな高頻度語を特別のルーチンで無視する方法である。もちろん get とか go というような若干の高頻度語はイデオムルーチンによって正しく同定出来る。あいまいではないが He thinks that……とか I feel that……とかの think と feel のような従属節を導く動詞としての用語が本動詞（例えば I feel energetic の feel など）としてのそれを比較して大きいカウントを示す場合は，従属節を導く用法を同定し，それを別個に扱うルーチンを用意する。

次に項目語の特性を最もよく表わすカテゴリーを付与する問題がある。例えば king は social role, high-status, political, male role 等の意味があり，いずれも king を定義するのに意味がある。したがって king にこれらの複数のカテゴリーを与えてよいかのいわゆる multiple tagging の問題である。その対策としては，一つにはカテゴリー間の相関を調べ，＋.80以上の場合重複測定（overlapping measure）が無意味であるから利用をとりやめ，単一のカテゴリーをあてる。Harvard Ⅲ辞典ではすでに利用者に重複カテゴリーの相関を知らせる方法を採用している。

辞典編成について，Woodhead らが試みつつある手続では，その段階を二つに分けている。第1段階では，理論に基づいて，どの種類のカテゴリーを用いるかを決定する。第2段階では，カテゴリーに属する語，イデオムを明細化する。新しい概念があらわれた場合，新しい項目を決定したり，他の辞典を参照することもこの中に含まれる。

第2段階の問題点は，第1はカテゴリー内の単語とイデオムの候補項目のリストを如何にして得るか，第2はどの項目をその中から選択するかという点である。General Inquirer の初期では，内容分析辞典は他の辞典を参照したり，委員会の討議を経て作った直感的産物であった。とくに同形異議語や，イデオムの項目決定は困難な問題を含み，現在でも慣用語辞典が用いられたり，構文編集やイデオムルーチンの技術を利用して意味を区別している。しかしいずれにせよ，きめ手がなく，語やイデオムの使用語あるいは頻度の記されているカタログがないということが致命的である。これらの問題を解決する第1歩として，KWIC が，とりわけ特殊辞

典の編集に有効に機能する。それは第1に項目語リストに頻度を提供し，第2に一般的イデオムとその出現頻度を同定し，第3に構文手続を組立てるのに重要な資料を提供するからである。

将来の問題としては，注目すべきものは source dictionary と呼ばれるもので，適切な KWIC 手続によって一般辞典，特殊辞典の両方の組立に利用出来る長期的に必要な資源の役割を果す。これは KWIC サンプルから5000の最も頻度の高い語を取出し，その一語一語について頻度情報を始め，慣用法，構文論，意味論等の観点から綿密な検討をなしたくわしい辞典である。これによって内容分析辞典編成の仕事を軽減し，内容分析の正確性を増すことが出来る。

6 むすび

以上，われわれは General Inquirer を中心に主として内容分析の技術面を概観して来たが，内容分析の情報検索，とくに文献検索とはかなり多くの共通の場をもっているといえよう。いずれもが情報の記録された媒体，つまりテキストを対象としており，内容分析におけるテキスト中の特殊な特徴を体系的・客観的に同定する作業は，まさに情報検索における主題分析のそれと類似している。もちろん，内容分析のこのプロセスは推論のためのものであるから，おのずから主題分析とは異なる要素をもつであろうことは否定できない。話したり，書いたりした人のあるトピックに対する態度の強度とか関心を測定するという意味では，それは政治的文書に代表されるような社会事象そのものに密着した記事の主題分析には有効に機能するであろう。また，対象が学術文献の場合，ある所説に対する筆者の態度の特定には有効であろうが，そこに存在する一般的な主題を特定するには若干問題なしとしない。

検索の面からみた場合には，General Inquirer は文単位の検索が中心で，文献単位のそれは直接には対象としない。したがって General Inquirer の諸技術とくに検索面のそれが文献検索にどの程度有効に機能するかは検討を要するところであろう。

内容分析の思想は，Luhn の自動索引あるいは KWIC に有力な示唆を与えたと考えられるし，現在でもまた，General Inquirer と情報検索の自動システムである SMART とは相互に影響し合いながら，それぞれの発展を遂げてきているといえよう。General Inquirer の諸技術の何がどのように情報検索に利用できるかを具体的により詳細に検討する必要があるが，このことについては紙幅の関係上他の機会にゆずることにして今回は以上 General Inquirer の単なる紹介にとどめた。