

情報メディアの構造: 伝達内容の分析と利用

Some Aspects of the Structure of Informational
Content of the Information Media

神 門 典 子
Noriko Kando

Résumé

This paper reviewed recent studies on some aspects of the content structure of the information media. The content of information media is not just a random grouping of words, but rather carefully constructed set of concepts put together by an author to express the author's idea. Therefore as a essential nature of the information media, its content involves structure. After types of structures and terms are defined, relations between structures and subject analysis, subject representations with relationship among concepts, structural analysis of abstracts and research papers, and information needs from structural point of view are discussed. Needs for validated analysis of structures and cumulation of studies are emphasized.

- I. 情報メディアの伝達内容と構造
- II. 構造の種類と用語の定義
- III. 構造分析の意義
- IV. 主題分析と情報メディアの構造
 - A. 主題分析の手がかりとしての構造
 - B. 索引作成の過程と情報メディアの構造
 - C. 抄録作成の過程と情報メディアの構造
 - D. 自動的な主題の分析と抄録
- V. 情報メディア内の概念間の関係の表現
- VI. 抄録の構造分析
 - A. 抄録の特性を示すための構造の分析
 - B. 抄録の機能構造の情報検索への応用

神門典子: 慶應義塾大学大学院文学研究科博士課程図書館・情報学専攻, 東京都港区三田 2-15-45
Noriko Kando: Graduate School of Library and Information Science, Keio University, 2-15-45, Mita,
Minato-ku, Tokyo.
1992年9月5日受付

VII. 全文を対象とした構造の分析

- A. 全文の特性としての構造
- B. 全文データベースの検索への応用

VIII. 情報ニーズの表現と構造についての知識

IX. 情報メディアの伝達内容の分析と利用という観点からみた今後の展望

- A. 図書館・情報学における情報メディアの構造研究
- B. 諸領域における構造分析の手法
- C. 情報メディアの利用からみた特性と今後の展望

I. 情報メディアの伝達内容と構造

図書館・情報学では、従来から、情報メディアによって伝達される内容には、何らかの構造があり、主題分析において重要な役割を果たしていることが漠然と認識されており、テキスト言語学や認知心理学の手法を応用した研究も次第に増えている。しかし、構造の捉え方はさまざまであり、用語も一定していない。

そこで、本稿では、最近 10 年間の図書館・情報学における情報メディアによって伝達される内容の構造に関する研究を検討すると共に、関連する諸領域での文章や談話の構造の捉え方やその基本的な分析法を概観し、「構造」の種類、分析法、構造を分析する意義を整理する。

ここで、情報メディアとは、人間の知的創造の成果をひとまとまりのものとして記録し、伝達するものを意味する。これは「文献」または「資料」とほぼ同義であるが、紙に限らずあらゆる物理的形態のものも含むものとする。本稿では、学術情報メディアを中心に考察する。

II. 構造の種類と用語の定義

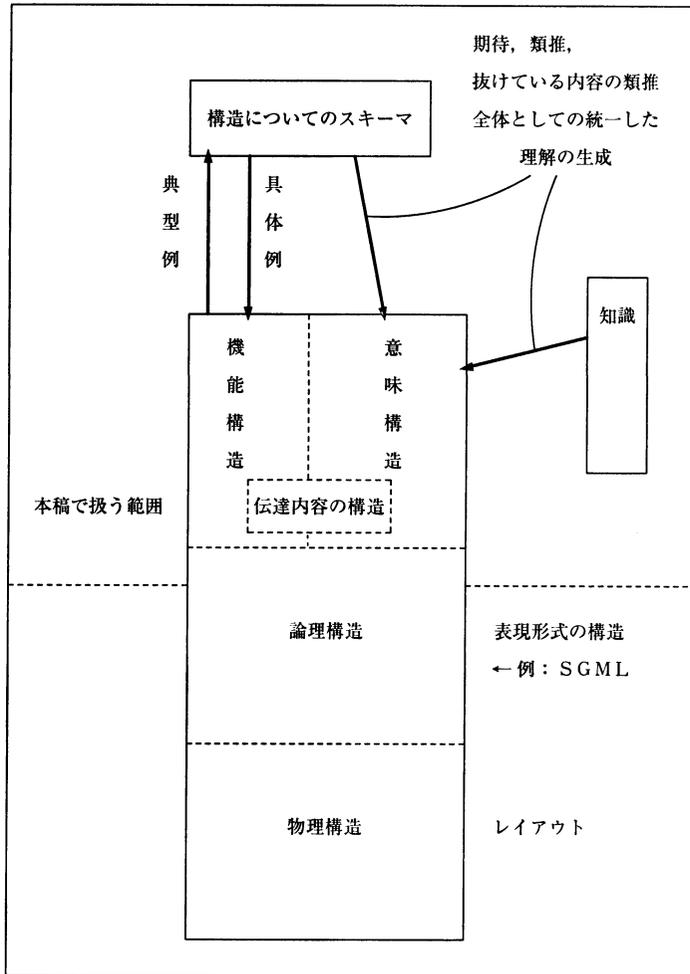
個々の情報メディアがもっている構造は、第 1 図のように階層化して捉えることができる。最も下位の物理構造は、物理的に媒体の上に固定されたところにみられる構造で、印刷されたページのレイアウトなどがこれに相当する。これから物理的な固定を取り除き、標題と本文、あるいは章、節、段落、文、章や節の見出しといった表現形式の区別を抽象的に捉えたものが論理構造である。Standard Generalized Markup Language (SGML) (ISO8879, 1986) はこの論理構造を記述する言語であり、論理構造を取り入れた検索方式も研究されている。

さらに、このような表現形式の区別も取り払って伝達されている内容そのものが持っている構造を考えることもできる。その一つは、伝えられている内容が、どのよ

うな要素からなり、それらが互いに関連しあい、全体としての意味を作り上げているさまを示すもので、意味構造といえる。それと同時に、伝えられている内容が、そのメディアの中でどのような役割や機能を果たしているか、すなわち、背景や目的、実験方法といったそれぞれの部分の機能に着目して構造を捉えることもできる。この機能面から捉えた枠組みとそれぞれの部分が持っている意味とが重なりあって全体の内容を作り上げていると考えられる。また、情報メディアによって伝えられる「情報」とは、物理的制約や表現形式の区別を取り払っていったこの伝達内容の本質的な部分にあると考える。

一方、認知心理学では、文章の理解には、文章以外の、読み手が持っているさまざまな知識が関わっており、その知識によって文章中の情報が結び付けられ、統一のとれた全体としての理解を組み立てることができる（内田, 1983）。その知識の特別なものとして文章の典型的な構成に関する知識がある。これは、たとえば、学術論文なら、背景、目的、方法、結果、考察という順で記述されるであろうというような文章の種類毎に一定の典型的あるいは一般的な構成に関する知識であり、これによって次にくる内容についての期待が喚起され、理解が促進するとされている。本稿ではこの文章の典型的な構成に関する知識を「構造についてのスキーマ」と呼ぶこととする。

また、この構造についてのスキーマは、文章全体からみた機能を表わすカテゴリーとその規範的な（あるいは可能な）順序から成る (Kintsch *et al.*, 1978)。したがって、これは個々の文章の機能構造の典型例として、その文章の作成や利用に関わる人に受け入れられているものと考えられる。さらに、文章自体の全体の意味を捉えていくとその上部で文章の構成についての知識に基づく構造と結びつくと思われる。この意味で機能面からみた全体的な典型的構造を van Dijk は「上部構造 (super-structures)」といっている (van Dijk, 1980)。



第1図 構造の種類と本稿の範囲

以上より、本稿では、情報メディアの構造として、伝達内容がもっている意味構造と機能構造、および伝達内容の理解に深く関わっているとされる「構造についてのスキーマ」を取り上げることとする。なお、立場によって「文章」「テキスト」「談話」という語が使い分けられるが、いずれも“文を素材・最小の単位体とする、より大きなまとまり”（寺村ほか、1990）という同等の意味とみなし、さらに情報メディアの伝達内容とも非常に近い意味とみなす。

III. 構造分析の意義

図書館・情報学では、従来から、情報を構造として捉

えたり (Belkin *et al.*, 1976)、文献の構造は主題分析に重要である (Wilson, 1978) とされている。

また、構造という語の一般的な意味は、「大辞林」(三省堂刊)では“さまざまな要素が相互に関連しあって作り上げている総体。また、各要素の相互関係”。あるいは、「日本大百科全書」(小学館刊)では“構造が構成されることの『要素』”, “構造が最終的に志向している『目的(性)』”, “構成要素の算術的加算を超える、構造の『全体性』”という三つの中心的な問題がある「概念」としている。

すなわち、構造を分析するということは、それを構成している個々の要素や要素間の関係に着目しつつ、それ

らの要素が関連しあってつくりあげている全体にも着目することであり、その全体は単に要素を羅列した以上の働きを持つものとして捉えることを意味している。

情報メディアの伝達内容も、単なる要素の羅列ではなく、要素間の相互作用によって全体として伝えている意味がある。その一方で、情報メディアの伝達内容は、利用者の目的に応じて特定の部分だけが利用されることもあり、部分部分にも意味や機能があると考えられる。また、従来から O'Conner (O'Conner, 1981) のように、利用者が必要とする部分だけを提供することによって、より直接的に利用者のニーズを満たすことができるという考え方もある。しかし、情報メディアのなかの部分部分は完全に独立した単位ではなく、そのメディアの文脈の中で妥当性や意味を捉えなければならない場合もある。このような情報メディアの特性から、その伝達内容を構造として捉えることは、全体の文脈の中で、部分部分の意味や働きにも着目していくことを意味し、より柔軟に捉える適切なアプローチであると思われる。

また、構造主義の立場では、『構造』とは、要素と要素間の関係とからなる全体であって、この関係は、一連の変形過程を通じて不変の特性を保持する” (大橋, 1979) と定義されている。ここでも要素、要素間の関係、全体が問題となっているが、さらに、それらの関係は表層が変化しても不変の特性を保持するとされ、これは構造としての情報の捉え方に通じるものであると思われる。

本稿では、要素、要素間の関係、全体という3点に着目して、情報メディアによって伝達されている内容の構造を捉えることとする。IV章以降、図書館・情報学における情報メディアの伝達内容の構造に関する研究を、情報メディアの特性および伝達内容の分析や情報検索への応用という観点から検討し、最後に関連領域における基本的な構造の捉え方を概観した上で、図書館・情報学における情報メディアの特性を踏まえ、今後を展望する。

IV. 主題分析と情報メディアの構造

A. 主題分析の手がかりとしての構造

図書館・情報学では、従来から、主題分析において情報メディアの伝達内容の構造の重要性が指摘されてきた。

たとえば、Patrick Wilson (Wilson, 1978) は、“文献は論理的・修辭学的関係をもって高度に構造化されているので、トピックの同定や要約には内容の構造の認識が

重要である”し、本文を通覧すると、“ある部分は歴史的背景、問題提起、その可能な解決、提案の問題点の吟味などを述べているということが判り、内容の構造を見いだすことができる。その内容の構造は階層がある木構造で表現することができ、その木構造の幹はその著作の目的を表わし、枝の末端の方から幹へとみていくとその著作の目的・戦略を読み取ることができる」としている。

また、Bernd Frohmann (Frohmann, 1990) は、図書館・情報学の理論的基盤の強化には、その主要な位置にある「主題分析」の理論的な解明が必要であるとしている。そして、索引すべき重要な語句や概念を同定する認知過程の研究は少なされているが、索引の質を高めるには実際に行われていることを認知的に記述するのではなく、具体的な重要概念・語句の同定法や指針、規則を明らかにする必要があると指摘し、それにはテキストの構造的な特性を用いることが有望であると述べている。

このテキストの構造的特性を用いると、強力な論理構造があるテキストでは、定理・仮説・前提・議論・結論・証左などの明らかな区別が示され、テキストの中心と周辺部あるいは主要なテーマと主要でないテーマとが明らかになり、修辭学的構造に依存しているテキストは、説得の手段と目的との明解な区別が示され、それによってテキストの中心と周辺部との差を明らかになると述べている。

すなわち、主題分析に重要な情報メディアの「構造」として、①歴史的背景、問題提起、その可能な解決、提案の問題点の吟味 (以上 Wilson)、定理、仮説、前提、議論、結論、証左、説得の手段と目的 (以上 Frohmann) など個々の部分が果たしている役割や機能に注目した「機能構造」と、②著作の目的を中心の幹とする木構造にし (Wilson)、中心と周辺部とを明らかにする (Frohmann) という全体的な「意味構造」との双方を意味している。また、Frohmann は、このような全体的な構造だけでなく、主語の繰り返し、指示語の使用や省略などの前方照応、文と文との間の論理的・修辭学的な関係などによって認定される数文の間の局所的な構造も重要であるとしている。

B. 索引作成の過程と情報メディアの構造

Clare Begthol (Begthol, 1986) や John F. A. Farrow (Farrow, 1991) は、テキスト言語学に認知的な側面を取り入れた後述する van Dijk の過程志向モデルに

基づいて、分類作業や索引作成作業の過程のモデルを提示している。

Farrow は、索引作成、分類、抄録作成を含めた主題分析の過程は以下の点で通常の文章理解と異なるとしている。すなわち、①作業時間が短い、②主題分析と主題表現作成という限定された目的のためだけに読む、③読んだ直後に主題表現を作成する、④特定領域で反復して作業を行なっている、という主題分析作業の特殊性を考慮し、語の出現頻度やテキスト構造を示す語を手がかりとし、主題・テキスト構造・索引システム・利用者に関する知識や一般的な世界知識を用いて選択的にテキストを読むというモデルを提示している。主題分析では、時間的制約の下で、特定種類のテキストで反復して作業することから、通常の文章読解の過程に比べ、テキスト構造のスキーマが特に重要な役割を果たしていると考えられる。

一方、実際の索引作成の認知過程を調査した研究として、大沼美佐(大沼, 1988)がプロトコル分析により、実際の索引作成過程では Anderson (Anderson, 1985)の「概念索引」と「シンボル索引」という二つのモデルに相当する処理がなされていることを示した研究などがあるが、選択された概念や語句とそのテキストでの位置やテキスト構造との関係には関心が払われてない。また、情報メディアと主題分析作業との関係については、Kevin P. Jones (Jones, 1983)は実際の索引作成作業の過程を調査し、索引する概念や語句は最初と最後の段落から多く抽出されることを示しているが、情報メディアの伝達内容の構造と索引作成作業との関係を示す実証的な研究はほとんど行われてなく、今後の課題のひとつといえるであろう。

C. 抄録作成の過程と情報メディアの構造

抄録作成の教科書や規格では、抄録に含めるべき要素を規定している。たとえば、「科学技術情報流通技術基準：抄録作成」(SIST-01, 1980)では、原著論文の場合、抄録に含める要素として (a) 前提 (b) 目的、主題範囲 (c) 方法 (d) 結果 (e) 考察、結論 (f) その他をあげ、(b) (c) (d) を詳しく書くとしている。E. T. Cremmins (Cremmins, 1982) は、抄録には (1) 目的、範囲、方法、(2) (1) の付加的な情報、(3) 結果、(4) 結論と勧告を含めるとし、Jennifer E. Rowley (Rowley, 1988) は、結果、結論、勧告、考察、今後の研究という見出しの節から重要な部分を引き出すとしている。見出しという表現形式

を手がかりとしているものもあるが、いずれも機能面からみた要素に着目しているといえるであろう。

一方、抄録作成の認知過程に関して、Farrow (Farrow, 1991) は、前述の索引作成過程モデルによるアバウトネス・モデル形成後、テキストの構造的手がかりに基づいてより詳細に処理が行われるというモデルを示している。

また、杉原寛子(杉原, 1985)は、実験によって、抄録作成作業における抄録作成者の認知過程を調べている。その際、抄録作成基準で考慮する項目として指示している前提、目的・主題範囲、方法、結果、考察・結論などを、抄録および論文の「マクロ構造」とみなしている。そして、抄録作成基準に関する知識がある抄録作成者はこの「マクロ構造」に関する知識があるとみなし、抄録作成基準に関する知識の有無と実際の抄録作成作業における認知過程との関わりを調べている。

なお、「マクロ構造」とは、一般に文章全体の概略、大枠、骨組みにあたるものを意味するが、杉原が指摘している「マクロ構造」に関する知識は、機能面からみた抄録や論文全体の構造に関する知識であり、本稿では抄録および論文の構造についての「スキーマ」としているものにあたる。ここでは用語を統一するために、このような機能面からみた全体の構造という意味で用いられている「マクロ構造」という用語を抄録や論文の構造についての「スキーマ」と呼ぶこととする。

杉原は、実験に当たり、抄録作成基準に関する知識がある被験者は、抄録作成基準に示されている抄録に含める項目についての知識としての「スキーマ」に従って論文から内容を抽出し、その内容を切り詰めていって抄録を作成すると仮定している。そして、この仮定の下に、抄録作成基準に関する知識の有無と、作成された抄録に含まれている内容や被験者が抄録対象となった論文の内容を理解した程度との関係を調べている。その結果、抄録作成基準の知識がある被験者はこの「スキーマ」に従って抄録に含める要素を抽出していることが示唆された。

さらに、杉原は、抄録対象となった論文を解析して、パラグラフを単位とした論文の「枝分かれ構造」を示している。前述の「スキーマ」が機能的側面を扱っているのに対し、この「枝分かれ構造」は意味構造に相当する。すなわち、「枝分かれ構造」では、一つの論文全体が意味的なまとまりを示す木構造として表現される。そこでは、枝分かれ水準の高いパラグラフほどその論文の中心的な

内容を述べており、枝分かれ水準の低いパラグラフはより特定の、あるいは論文全体から見たら瑣末な内容を述べていると考えられる。

そして、抄録作成者が原文の内容を理解した程度、および、原文から抄録へ情報が抽出された程度をパラグラフ単位に調べ、各パラグラフの枝分かれ水準との関係を分析している。その結果、枝分かれ水準が低いパラグラフほど抄録作成者の理解度が低い傾向がみられた。また、論文から抄録へ情報が抽出されている程度に関しては、第3水準以上に位置付けられているパラグラフ、すなわち比較的中心的な内容を述べているとされているパラグラフの内容は抄録に抽出される程度が高かった。それに対し、それよりも低い水準、すなわちあまり中心でない内容を述べているとされるパラグラフでは、そのパラグラフの内容が抄録に抽出される比率が著しく低い傾向がみられたとしている。

しかし、提示されているデータを検討すると、被験者の抄録作成基準についての知識の有無によって、同じ第5水準に位置付けられているパラグラフ、すなわちあまり中心でない内容を述べているとされているパラグラフでも、そのパラグラフの内容が論文の中で果している役割によって論文から抄録へ情報が抽出される程度に異なる傾向がみられた。すなわち、「過去の実験」について述べているパラグラフに比べ、「著者の実験の手順」について述べているパラグラフは、抽出度が高い(50~100%)傾向が認められる。この「著者の実験の手順」は、抄録作成基準において、抄録に必ず含め、詳しく書くとされている「方法」に相当する。方法に関するこれらのパラグラフから抄録への情報抽出は作成基準に関する知識の有無による差が最も顕著に現れた点であった。したがって、この杉原のデータから、作成基準に関する知識がある被験者は、「枝分かれ構造」という意味構造で表わされるような中心的内容かどうかという区別より、「スキーマ」とその要素に相当する特定の「機能」を果たしているパラグラフかどうかということが、抄録への情報抽出へ大きな影響を与える場合があることが示されたと考えられる。実際の抄録作成過程を、情報メディアの機能構造および意味構造と構造についてのスキーマとの関係で分析した興味深い研究であると考えられる。

一方、武者小路澄子(武者小路, 1988)は、情報メディアの伝達内容がより濃縮された抄録というメディアとして生成される過程に注目し、論文の本文と抄録とを比較し、抄録作成の過程に生じる操作と内容や表現の変化

を質的に検討している。Begthol (Begthol, 1986) が指摘しているように、理論的には、論文の本文と抄録とは同じアバウトネスを示すはずである。しかし、武者小路は、実際には抄録作成過程でさまざまな変形操作が生じていることを示し、論文の伝達内容や表現形式との比較という観点から「抄録」というメディアとその形成過程の特性を捉えている。さらに、研究者へのインタビューとその作成した著者抄録との比較から、抄録作成時には一般的な構造についてのスキーマだけでなく、個人個人が書くべきメディアについて個別に持っている詳細なスタイルのイメージに従って内容を構成していることを示している(武者小路, 1989)。これらの武者小路の研究は、情報メディアの内容を理解して分析する処理としての主題分析ではなく、むしろ抄録というメディアを作成する過程に着目しているという点で特徴的な立場をとっている。

D. 自動的な主題の分析と抄録

Udo Hahn (Hahn, 1990) は、情報メディアの伝達内容を自動的に分析し、中心となるトピックを同定するシステムの研究をしている。そのアプローチは、まず、語句の前方照応や省略を手がかりとして繰り返し主語として出現する概念を明らかにし、同じことがらについて述べている局所的なまとまりを同定する。そして、知識ベースを用いてそれらの繰り返し出現する概念の意味や概念間の関係や局所的なまとまりの間の関係を理解することによって自動的にテキスト全体の意味構造を分析し、中心となるトピックを同定するものである。

それに対し、Chris D. Paice (Paice, 1990) は、自動抄録に関して、Hahn のような人工知能的手法は領域の知識に関する知識ベースが必要であるため、限定された領域でしか適用できず、近い将来、論文全般に適用できる見込みがないとしている。そして、語・出現位置・構文などを手がかりとして、論文から重要な文を抽出し、それらの抽出文間の前方照応関係を同定し、意味が通るように加工して抄録にするという実務的なアプローチの方が有望であるという観点に立ち、1950年代後半から1970年代の実務的なアプローチによる諸研究をレビューしている。そして、よりバランスのとれた抄録を作成するには、「重要な文」というあいまいな基準よりも、領域を超えて共通性がある機能面からみたテキスト構造が、文を抽出する枠組みとして有望であるとし、この実務的なアプローチでの自動抄録の発展は、適切なテキスト構

造理論の発展が不可欠であると結論づけている。

ここで、前方照応関係は、①局所的な意味のまとまりを捉える手がかり、②抽出文の意味的な完結性、③語の出現頻度の計数の前処理、として情報メディアの伝達内容の構造に関わりがある。

①は、同じ語が繰り返し出現する部分は同じことについて述べられている意味的にまとまりある部分と考えられる。多くの場合、繰り返される語は、意味的な一つのまとまりの範囲内では指示代名詞などの照応表現に置き換えられている。したがって、このような照応表現をその被指示語が繰り返し出現したものとみなすことによって局所的な内容のまとまりを認定することができる。

②は抽出文中に非抽出文を指示した照応表現が含まれないように、照応表現を被指示語に置き換える、あるいは、被指示語を含む文まで抽出することである。このような観点では、Timothy Craven (Craven, 1988) は、個々の利用者に適した抄録を個別に提供することを目的として、抄録の一部分だけを取り出しても、意味の通るものとなるように、前方照応などの文の依存関係を調べている。また、抄録全体を文間の依存関係を含めて圧縮した形でディスプレイ上に提示し、対話型で個々の利用者の必要な部分を詳しく提示する方法を検討し、実験システムを構築している (Craven, 1990)。

③では、重要な概念ほど言及回数が多く、指示語や代名詞などで置き換えられる場合が多いと考えられることから、正確に語の出現頻度を計数するには照応表現を被指示語に置換する必要がある。このような問題意識の下に、シラキュース大学の研究グループは、自然語検索において文を超えたレベルの事象を扱う研究の一環として、複数領域の抄録から前方照応の用例を集め、1抄録あたりの前方照応数を調べ、前方照応を判定するルールを検証している (Liddy *et al.*, 1987; Bonzi *et al.*, 1989)。

以上、主題分析と情報メディアの構造との関係を概観した。主題分析の理論的基盤や具体的な規則の基盤として情報メディアの伝達内容自身も持っている構造の有用性が示唆されている。すなわち、中心的な内容を述べている部分に着目し、述べられている概念間の関係を把握するという点で意味構造が主題分析の過程に関わっていると考えられる。また、主題を捉える上で重要な特定の役割を果たしている部分に着目し、全体をバランスよく捉えるという点で機能構造が関わっていると考えられる。さらに、その基礎には、語の繰り返しや省略、文と

文との関係などから把握できる局所的な構造が関わっていると考えられる。主題分析は、認知的モデルとして通常の文章理解よりも制約された過程と位置づけられ、構造に関するスキーマが強く働くと想定され、実際の抄録作成作業に関する研究においてもそれが示唆されている。しかし、実際に主題分析と情報メディアの伝達内容自体の構造との関係を分析した研究は少なく、今後の課題の一つであろう。

また、主題分析の結果として作成される主題表現において、索引語には多くの場合、情報メディアの伝達内容が持っている要素間の関係や構造は反映されない。そこで、次に、情報メディアの伝達内容の中にみられる概念間の関係の表現に関する研究を概観する。

V. 情報メディア内の概念間の関係の表現

図書館・情報学では、従来から、情報メディアによって伝達される内容、すなわち、情報はバラバラの事実ではなく、基本的な特性として要素間の関係を含む「構造」があるという考え方があった。たとえば、Nicholas J. Belkin と Stephen E. Robertson (Belkin *et al.*, 1976) は、構造としての情報という概念を種々の複雑さの程度をもつスペクトラムとして提示した上で、情報学においては、情報は、「受け手のイメージ構造を変化させることができる何らかのテキストの構造」であるとしている。Eliahu Hoffmann (Hoffmann, 1980) も、情報は、事実や数値とそれらを結び付ける意味のある関係であると定義している。

一方、索引語は多くの場合、主要な概念を表現することに集中しており、それぞれの情報メディアの伝達内容の中で関係付けられている概念間の関係や構造は表現されていない。これは、情報メディアの伝達内容の本質を十分に表現しているとはいえない。

そのなかで、リンクやロール・インディケータは、概念間の関係を表現しようとした試みの一つである。また、Farradene (Farradene, 1980) は、主題概念の多様性に比べ、概念間の関係は限定され、領域を越えた共通性があることに着目し、文献中の概念間の9種類の論理的関係を表わすことのできる関係索引法 (relational indexing) を提唱し、Weiner (Weiner, 1982) は医学論文における数値間の関係の重要性に着目した索引法を提唱している。

さらに、Sherrilynne Fuller (Fuller, 1984) は、スキーマ理論を応用して文献中の概念間の関係を表現するこ

とを試みている。これは、主題 (aboutness) や重要な部分 (importance) だけでなく、主要でない部分でも利用者には必要な場合もあるという問題意識の下に、次のような要件をみたす索引法として提案している。すなわち、①主観性の排除、②一貫性の向上、③事後結合による誤結合 (false-drop) 防止、④当該領域で研究者間に共有の知識体系に従う、⑤文献内の概念関係の表現、⑥文献間の概念関係の表現、⑦論文の質を評価する手がかりとなる可能性がある、という要件を満たしている。

そして、医学における臨床試験報告論文をとりあげ、共通する概念カテゴリを抽出し、図式化して「臨床試験報告のスキーマ」とした。これは、この領域の研究者間で共有されている知識体系に従い、文献内の概念の関係を表わし、文献間で共通の概念の枠組みとなるものである。MEDLINE から無作為抽出した臨床試験報告論文 482 件を対象とし、このスキーマを用いて索引作業を行っている。各論文から、スキーマの各スロットに当てはまる概念を抽出し、あるいは、その要素の有無を答える形をとったことにより、主観性が低減されたとしている。また、複数分析者による同一文献の索引結果は一致度が高く、索引作成作業の一貫性が示された。さらに、臨床試験報告は人命に関わる医薬品についての重要な情報源であり、個体差の大きい人間を対象としていることから研究デザインへの関心が強い。それを踏まえ、この Fuller のスキーマでは盲検や統制群の設定など研究デザインの質に関するスロットを設けていることから、論文の質を評価する手がかりにもなる。以上より、この索引法は上記の要件を満たすと結論付けている。

Weiner や Farradene は領域を越えて共通する概念間の関係を扱っているのに対し、Fuller は医学の臨床試験報告に限定したことにより、より詳しいスキーマを設定することができた。これは、限定された領域における知識体系や規範的な研究方法および研究者が共有している情報メディアの伝達内容の構造に関する知識に従って、概念を抽出するとともに概念間の関係も表現するものである。このように情報検索では、情報メディア内の概念間の関係を個別的に表現するだけでなく、それを情報メディアの集合の中で、共通の枠組みで捉える必要がある。

しかし、このように情報メディアの伝達内容の中の概念間の関係や構造の表現を実用化するには、作業負担だけでなく、一貫性や妥当性、検索における効果についても検討する必要があるだろう。Fuller のスキーマ理論

を応用した索引法では、論文中での各要素の出現位置や手がかり語句が比較的一定であることから、自動化も可能であろうと述べられている。作業負担を軽減し、分析基準を明確にする上からも自動化は望ましいと思われる。

VI. 抄録の構造分析

A. 抄録の特性を示すための構造的分析

抄録の機能構造を分析した研究として以下のものがある。A. B. Buxton ら (Buxton *et al.*, 1978) は、本文中の緒言・方法・結果・考察の各節とそれに相当する抄録の内容との語数や関係を調べている。Takashi Maeda (Maeda, 1980) は抄録の機能構造の自動分析の一環として、Milica Milas-Blacovic ら (Milas-Bracovic *et al.*, 1989) はユーゴスラビアにおける抄録の質の評価のために、Helen R. Tibbo (Tibbo, 1992) は歴史学文献に適した抄録の構造と内容を明らかにする研究の一環として、それぞれ、さまざまな領域の実際の抄録の構造を分析している。分析用のカテゴリはいずれも抄録作成基準に示された緒言 (または、背景、目的・範囲、仮説 (Tibbo, 1992)), 方法、結果、考察という 4 つもしくは 6 つであり、その有無や出現順によって分析している。

これらの研究は、いずれも共通のラベルをもったカテゴリを用いている。Tibbo は付録として分析方針とカテゴリの定義を示し、分析者間の分析一致度を信頼性として示している。また、Buxton らは分析例を若干示しているが、それ以外の研究はいずれも、カテゴリの定義や意味する範囲、分析基準、分析例が示されてなく、どのような文がどのカテゴリに割り振られているかは明らかではなく、分析の妥当性に関する記述はない。

B. 抄録の機能構造の情報検索への応用

抄録は、情報メディアの伝達内容の概要を示すだけでなく、多くのオンライン・データベースでは抄録中の語から検索できることから、索引語と同様に、主題からの検索の手がかりとしての機能があり、主題表現のひとつとみなすことができる (Lancaster, 1991)。前述の Tibbo も人文科学での自然語検索における抄録の重要性を認識し、歴史領域の抄録には標準的な内容要素に当てはまらない文の比率が他領域より高いという調査結果に基づき領域に適した抄録作成方針の策定の必要性を提案している。また、Raya Fidel (Fidel, 1986) も抄録は自然語検索に適した内容と構成を含むべきであるという

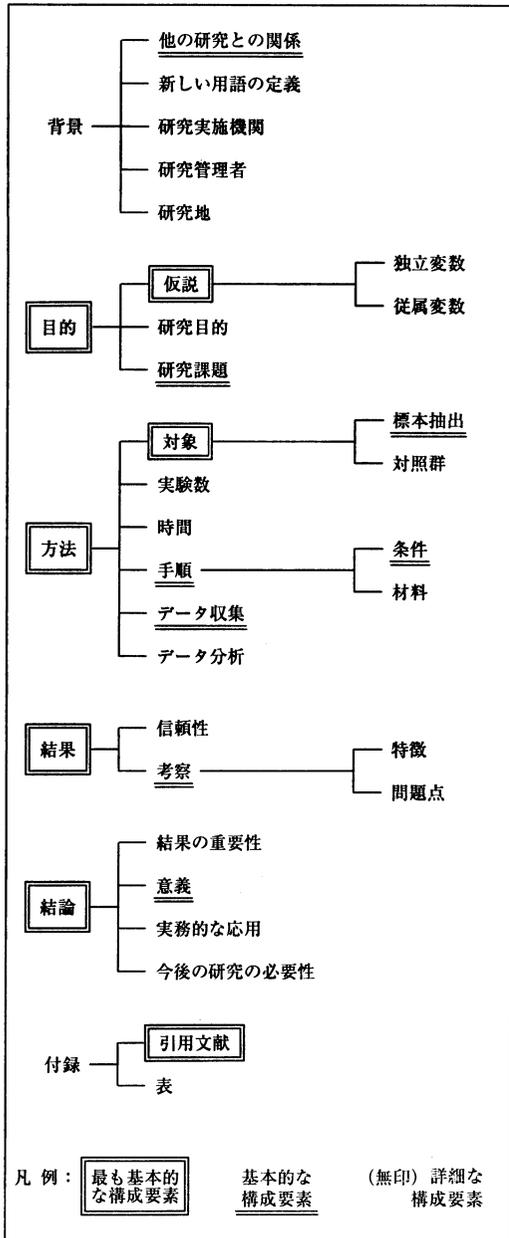
問題意識の下に、データベース作成機関の抄録作成方針を調査している。これらは現状の検索方式の範囲において、より適切な内容を含む抄録の作成を目指すものである。

それに対し、Liddy (Liddy, 1991) は、現在の情報検索システムは語を単位とし、その語が表わしている概念が文献中で果たしている役割や機能、および、概念間の関係を表現できないことが問題であるとしている。このような問題意識の下に、文献の内容のよりよい記述を検討する一環として、抄録中の文や語句が文献中で果たしている役割や概念間の関係を表現する方法を考案した。これは、抄録中の語句と、それらが果たしている役割とを組み合わせることで検索することにより、検索精度の向上が期待されるというアイディアに基づいている。

具体的には、談話言語学と認知科学のスキーマ理論から示唆を得て、実証的な研究を報告した論文の抄録を対象として機能構造を分析している。まず、抄録作成者の抄録構造に関する知識を聴取し、第2図のような抄録の構造のモデルを設定した。これは、抄録の内容を機能面からみた構成要素と各要素が果たしている役割という要素間の関係から成り立ち、「緒言—方法—結果—考察」という規範的な構造より詳細なものである。また、これは、抄録作成者が抄録の構造に対して持っている知識に基づくスキーマと考えられ、抄録の機能構造を分析する共通の枠組みともなる。

そして、教育学と心理学の論文の抄録それぞれ150件と126件に人手でこのカテゴリを付与して構造を分析した。また、そのうち、68件に関しては専門家である抄録作成者が分析し、比較評価を経た適切な構造と比較したところ、86%の精度で一致し、分析の妥当性が示されている。さらに、各カテゴリごとに比較的少数の特徴的な手がかり語が認定されたとし、これを用いて自動分析が可能であると展望している。

さらに、これをうけて、Robert Oddy と Liddy ら (Oddy *et al.*, 1992) は、この Liddy と同一の抄録を対象として、確率モデルとニューラル・ネットワークモデルを用いてカテゴリ付与の自動化を試みている。結果として実用化にいたる精度は得られなかったが、改善点が示唆され、有望であると結論付けている。それとともに、この Liddy が提案した抄録の機能構造を、概念間の関係や文脈・状況を領域を超えて共通に表わす枠組みと位置づけ、他方、検索質問においてもトピックだけでなく利用者の問題や状況を表現する必要があるとし、研究過程中的各段階を利用者の状況を表わすものとして取り込み、トピックや語の意味だけでなく、その文脈もしくは状況を扱うことのできる検索システムの必要性を提唱している。



第2図 実証的な研究の抄録の構造 (Liddy, 1991, p. 71 より翻訳)

これらの Liddy らの研究は、抄録中に表わされている概念間の関係を抄録のテキスト・データに付加することによって、より付加価値の高い表現をもったデータとし、検索の高度化をめざしたものと見える。

VII. 全文を対象とした構造の分析

A. 全文の特性としての構造

学術論文を例にとると、一般に自然科学領域では「緒言—方法—結果—考察」という規範的な構造があるといわれ、多くの論文執筆マニュアルで示唆されている。実際にこのような節の見出しを持つ論文も多く、一般に「テキスト構造についてのスキーマ」といった場合はこれを意味することが多い。

また、Milas-Bracovic ら (Milas-Bracovic *et al.*, 1989) は、このような節見出しがなくても該当する内容が述べられているとし、Buxton ら (Buxton *et al.*, 1978) は、他の節の方がふさわしい内容が含まれることがあるとしている。このように、実際の論文では、節の見出しに表われていなくても内容にはなんらかの構造があり、「緒言—方法—結果—考察」より複雑な現われ方をすることが想定され、分析もなされている。

たとえば、Jiri Janos (Janos, 1979) は、文の機能分析におけるテーマとレーマという概念を拡大し、テキスト中のテーマである主題、拡大テーマである背景とアプローチ、レーマである結論、拡大レーマである仮説、方法、報告というカテゴリを設定し、それによる機能構造分析を提案している。また、W. Dea と N. J. Belkin (Dea *et al.*, 1978) は、対比、言い替え、原因と結果、手段と結果、一般と特殊などの文間の関係を規定し、それぞれ関係を示す手がかり語を参考に、実際の論文において文の役割や文間の論理的・修辭学的関係を分析している。さらに、このように分析した文間の関係に基づいて、状況・問題・解決・観察と評価という論文全体を構成している一般的なメタ構造を分析し、重要な文の認定をしている。

さらに、全文全体を分けるのではなく、メディアの種類に特徴的な典型的要素を選定しその有無を調べた研究ではあるが、機能面に着目したのものとしては、P. L. Hibberd と A. J. Meadows (Hibberd *et al.*, 1978) がある。これは、臨床試験報告論文の特性を明らかにし、その内容を評価するために、6誌から選択した計 259 論文を対象として、患者の性別・年齢・人種・診断名・投与機関・経路・量、医薬品の製造者・剤形などの 35 種の

要素が含まれているかどうか調査し、専門領域別、年次別、および、雑誌の貸出頻度別に比較して特性を明らかにしている。

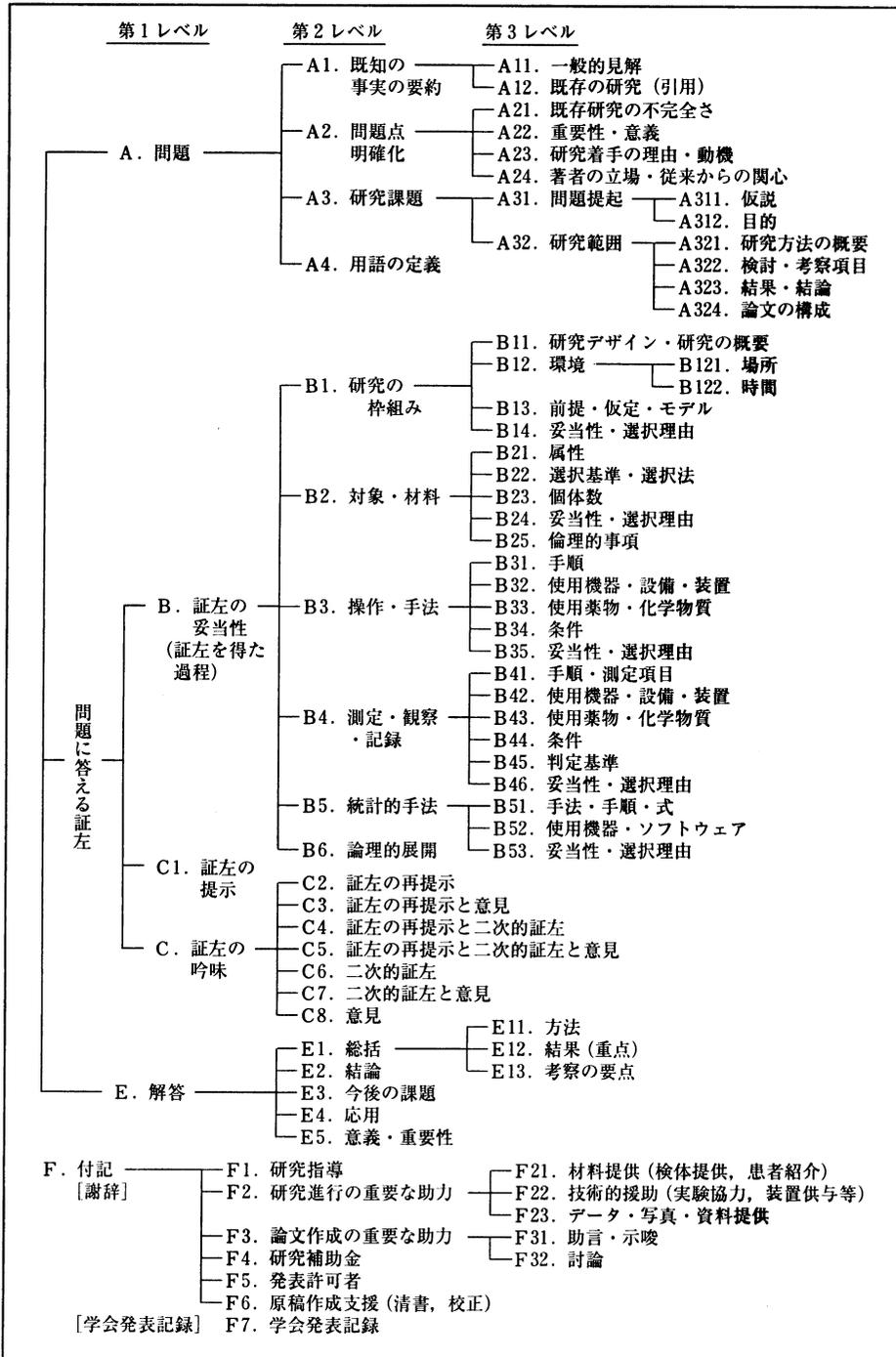
それに対し、神門典子 (神門, 1991) は、情報メディアの伝達内容の特性を明らかにするため、日本語の複数領域の原著論文を例として、その機能構造を分析している。情報メディア内部の関係を示し、情報メディア間で共通の枠組みが重要であるという観点から、意味構造は個々のメディアの個別の構造であり、要約や主題分析には関係があるが、メディア間で共通の枠組みとはならないことから、機能構造を選択している。

そして、論文執筆マニュアル 41 点の分析と実際の原著論文の分析を通して、第 3 図のような一連の「構成要素カテゴリ」を設定し、それを用いて原著論文の内部構造を分析した。このカテゴリは、情報メディアの伝達内容を領域や研究方法に関わらずに捉える共通の枠組みであり、各部分が論文中で果たしている機能を示し、部分と全体の関係と部分間の関係を示すものである。分析にあたってはカテゴリの定義と手がかり語などの分析基準に基づき、分析対象となる原著論文の各文に、人手で最も下位レベルのカテゴリを付与し、論文全体の構造を見る際には順次上位のカテゴリでまとめ、カテゴリの出現順によって構造を記述した。

その結果、分析対象とした医学、物理学、経済学、国文学の原著論文 127 件について、全ての文に構成要素カテゴリを付与することができた。カテゴリの出現順からみた論文の機能構造は、最も上位レベルでは共通性があり、部分的な下位構造の組合せ、省略、繰り返しを考慮すると共通性がみられた。また、実際の論文の機能構造は「緒言—方法—結果—考察」より複雑なものであった。全文データベースの検索や検索結果の提示などさまざまな応用が可能であり、また自動付与の可能性と必要性を示唆している。しかし、一定の分析基準に基づき、言語学的特徴も手がかりとし、分析例も示しているが、一人の分析者が分析を行った結果であることから、妥当性の検証は充分ではないとし、今後の課題として、複数分析者間での分析結果の比較やより多くの領域への適用を通じて、妥当性を明らかにしていく必要があるとしている。

B. 全文データベースの検索への応用

神門 (神門, 1991) は構成要素カテゴリを付与した全文データがあると仮定した場合、全文データベースの検



第3図 構成要素カテゴリ

索において語句とカテゴリを組み合わせるにより、検索効率を向上させ、特定の機能を持った部分だけを対象とした検索やカテゴリを用いて検索結果の必要な部分だけを提示したり、必要に応じてカテゴリ間の関係に基づいて関連のある部分を柔軟に提示することが可能であろうと述べている。Liddy (Liddy, 1991) や Oddy ら (Oddy *et al.*, 1992) は、現在のシステムが主として抄録を扱っていることから抄録を対象として、役割や機能を表わすカテゴリを組み合わせて検索するシステムの必要性を提案しているが、情報メディアの全文を対象とした分析も可能であると考えられる。

たとえば、Joost G. Kircz (Kircz, 1991) は、学術論文に含まれる知識を意味的な側面からのみ表現してきた従来の索引や情報検索研究の限界を指摘し、それを補う根本的に異なるアプローチとして論文の議論構造を分析することを提案している。個々の文が果たしている「理論的な前提」、「他への引用」「実験の制約」などといった議論構造における役割の分析は、意味的な面から表現する従来の索引作成作業に比べ、主題知識がほとんど必要なく、あいまいさも減少するとしている。しかも、領域を超えて共通の枠組みで捉えることができる。また、議論構造を分析するための試験的なテンプレートを示しているが、Kircz 自身述べているように、これは完全なものではなく全文データベースにおける議論構造分析の研究の出発点としてアイデアを提示したにすぎない。しかしながら、全文データベースにおいて、論文中の各文に、議論構造という観点からみた機能を表わすカテゴリを付与することによって、意味とは別の観点から論文中で述べられている要素の性質を規定し、要素間の関係を明らかにするという点で興味深い提案である。

前述のスキーマ理論を索引に応用した Fuller (Fuller 1984) は、情報メディアそのものではなくメディアから抽出した概念を扱っているという差はあるが、Fuller も、抄録や全文の機能構造を扱っている Liddy (Liddy, 1991)、神門 (神門, 1991)、Kircz (Kircz, 1991) も、①主題や重要な概念だけでなく、②語や概念が持つ意味だけでなく、③情報メディアの伝達内容の中にみられる要素間および全体と要素の関係を保持し、④伝達内容を捉える情報メディア間で共通の枠組みとなるという特徴がある。しかし、Fuller は情報メディア単位の検索を前提として要素を抽出して主題表現を作成しているのに対し、Liddy や神門は、テキスト・データそのものに役割表示を付与し、より付加価値の高いデータを作成するこ

とを意図している。さらに、全文を扱う神門や Kircz では情報メディアの部分を対象とした検索や表示をも考慮している点で特徴的である。

VIII. 情報ニーズの表現と構造についての知識

文献の構造についてのスキーマは、一定の種類 of 文献についてその領域の専門家のあいだで共通のものとされている。また、既存の研究において、文献の構造についてのスキーマを有している人は、スキーマの無い人よりも文献の内容をよりよく記憶し、より多く想起することが示されている。これらのことから、この構造についてのスキーマをレファレンス・インタビューに応用し、スキーマに従って、利用者が必要とする文献の内容についての具体的な情報を引き出す試みがなされている。

Bryce Allen は、レファレンス・インタビューにおいて、「文献の構造に基づく質問 (structural question)」が、文献の標題やキーワードだけをたずねる「書誌的な質問」や自由な回答を求める「オープン・クエスチョン」より有効であるかどうかを検討している。この「文献の構造に基づく質問」では、利用者が求めている文献の内容を、Kintsch ら (Kintsch *et al.*, 1978) が提示した目的・方法・結果・結論という「科学論文の構造についてのスキーマ」にそって利用者にとずねる。文献の構造についてのスキーマがあると、従来、文献の内容をよりよく理解したり記憶したりできるとされているが、それだけでなく、文献探索において、求めている文献の内容をよりよく表明することができるという仮定に基づいた研究である。

その結果、「文献の構造に基づく質問」は、他の質問より、多くのキーワードを利用者から引き出すことができ、有効であると考えられた (Allen, 1988; Allen, 1989a; Allen, 1990a)。また、文献の構造が一定しているトピックだけでなく、一定していないトピックに関しても有効であった (Allen, 1988)。しかし、情報検索についての知識がある被験者から引き出すことができたキーワードの数は、「オープン・クエスチョン」と「書誌的な質問」とで有意差がなかった。これは、情報検索では「書誌的な質問」に答えればよいという先入観が働いたと考えられる (Allen, 1988)。それに対し、利用者に情報検索を前提としているインタビューであることを告げなかった場合は、「オープン・クエスチョン」と「文献の構造に基づく質問」は、共に、多くのキーワードを利用者から引き出すことができた。しかし、「オープン・クエスチ

ョン」に対する答えには文献の標題や抄録には含まれない語や命題が、「構造に基づく質問」に対する答えより、多く含まれていた。このことから「構造に基づく質問」が特に有効であるとしている (Allen, 1989a)。

また、このような質問の型は、利用者の学問的背景 (Allen, 1990a) と複雑に影響しあい、利用者が知識を構造化する方法 (Allen, 1990b) ともなんらかの関連がある場合もあることが示唆されている。

これら一連の研究では、質問型による効果だけでなく、実験条件や利用者側の種々の特性が複雑に関係しあい、それぞれ限定的な結論を述べるにとどまっている。また、現状の検索システムには文献の構造に対応した検索機能がなく、したがって、Allen も指摘しているように、より多く表明されたニーズを生かし、現在のシステムで有効に用いるには仲介者が必要である。さらに、情報検索システム側にも、文献の構造の手がかりを生かす機能が備われば、検索の面での有用性の検討が進むと考えられる。Oddy ら (Oddy *et al.*, 1992) は、文献の構造の手がかりを情報検索システムと利用者側の状況との両面から検討を加えたものと位置づけられるであろう。

IX. 情報メディアの伝達内容の分析と利用 という観点からみた今後の展望

A. 図書館・情報学における情報メディアの構造研究

図書館・情報学における情報メディアの構造に関する研究を概観したところ、情報メディアの伝達内容を分析するにはさまざまな構造や構造に関する知識が関わっていることが示唆され、また、実際に抄録や全文の構造がさまざまな方法で分析されている。

機能構造は、メディアの種類ごとに共通の枠組みで内容を捉え、分析結果はわかりやすい。分析の目的によって、分析の詳しさが異なり、緒言—方法—結果—考察という簡潔な構造で充分とする立場もあるし、Liddy (Liddy, 1991) のように詳細なカテゴリを必要とする立場もある。領域や研究方法を限定すれば、Fuller (Fuller, 1984) や Hibberd ら (Hibberd *et al.*, 1978) のように詳しい要素へ展開することも可能であり、特定要素の有無や要素間の整合性から伝達内容の質の評価にも応用できる。その一方で、カテゴリの意味範囲、定義、分析基準が明確でない場合が多く、分析やカテゴリの妥当性を示すことが重要である。実用的な応用には、作業負荷を減らし、分析基準を精緻化するという点からも分析の自動化が必要である。また機能構造は、メディアの種

類に応じて分析の枠組みを設定する必要がある。

それに対し、意味構造では局所的な意味のまとまりの間の論理的関係の認定や木構造状に表現した際の水準の設定は一定ではなく、分析者により差が生じるともいわれている。また、個々の情報メディアの内容を個別的分析し、分析の際には内容についての知識が必要であるが、同じ分析手法で、異なる種類のメディアの内容の構造を分析することが可能である。

この機能構造と意味構造とは、全く別個のものではなく、意味構造の上部には機能構造が関わり、機能構造を分析するにも局所的な意味のまとまりを考慮するというように互いに補い合い、重なりあって情報メディアの伝達内容を作り上げているものである。

また、情報メディアの構造の重要性が従来から指摘されていたが、実際に分析を行った研究は少なく、十分な研究の累積がない。これは分析法や用語が安定していないことも一因であると考えられる。たとえば「テキスト言語学的分析」は文を超えたレベルにみられる事象を分析するあらゆるアプローチを意味するが、あたかも特定の分析手法を示すように用いられることもある。

一方、情報メディアの伝達内容の構造を分析することは、メディアを構成している部分部分に着目しつつも、部分間の関係や全体と部分の関係、総体としての全体にも目を向け、文脈を保持しながら部分を扱うことを可能にする。また、部分によって重要性や役割が異なることを明らかにし、情報メディアの伝達内容中の各部分の均質性を否定する。さらに、文と文との関係として人間が読み取っているものを表現しようとする試みであり、情報メディアの伝達内容の本質である「情報」へ向かうアプローチのひとつと位置づけられるであろう。

情報メディアの構造に関する研究の流れをみると、主題分析や主題表現の作成における手がかりとして個々のメディア毎に個別な構造を分析することから、構造を情報検索や情報メディアの伝達内容の多面的な利用を可能にする共通の枠組みとして用いる方向が示唆される。また、重要な部分を認定するための構造分析から、メディア内の関係を表現し、それをういたメディアの有効利用を志向する方向が示唆される。さらに、概念を抽出してその間の関係や構造を捉えることから、テキスト・データそのものに関係や構造を示すものを付加するという方向もみられる。これは、電子化されたテキスト・データの蓄積の量的な増大ともなう動きであり、情報メディアの伝達内容の特性を示し、情報メディア研究の基盤

を提供するとともに、全文データベースを改善する一つの方向にもなりうると思う。

B. 諸領域における構造分析の手法

前述のごとく、実際に情報メディアの構造を分析した研究は少なく、分析法や用語も一定でない。構造の分析については、テキスト言語学や認知心理学の手法を援用したものが多かった。そこで、これらの領域を中心にその発展過程と基本的な分析法を概観する。

1. 意味構造の分析

テキスト言語学は、de Beaugrande ら (de Beaugrande *et al.*, 1981) によると、文を超える単位であるテキストを対象とし、単一の理論や方法としてはまとまらないが、その多様さによってテキストの理解が深まる。また、テキスト言語学では、1960年代からテキストがひとまとまりと認定できる対象であることを示すために構造が分析されるようになった。これは結束構造（または統括構造、coherence structure）といわれ、語の繰り返し、文や句の接続関係や論理的関係に基づいて、テキストの内容の意味的なまとまりを示すために分析される。中心的内容と周辺的内容とを区別する働きもある。これは本稿の「意味構造」に相当する。

一方、心理学の文章理解研究では、Bonnie J. F. Meyer (Meyer, 1985a) によると、テキストに表現された書き手の潜在的な論理とイメージを調べる手段としてテキスト構造を分析し、それと読み手のプロトコルとを比較することによって、書き手と読み手の潜在的な論理やイメージの関係を調べる。したがって、ここではテキスト構造は研究の目標ではなく、書き手の論理やイメージがどの程度読み手に伝わったかを評価する基準として用いられ、研究目的に応じて多様な分析法が用いられている。これらの分析法は、それぞれ、分析された構造における階層的な水準の設定、テキスト以外の知識の使用、要素間の関係の規定などの点で特徴がある。

たとえば、Meyer (Meyer, 1985a) は、同一テキストの構造を異なる方法で分析している。それによると、Kintsch の方法では、テキストの内容を命題に書き直し、命題の集合であるテキストベースを作成し、命題間での変項の繰り返しを手がかりとして階層化した表現が作成される。それに対し、Meyer の方法では、テキスト中の概念ユニット間にみられるグループ化、原因、問題と解決、対比などの論理的関係に基づいてテキスト全

体の構造が分析され、階層化した表現が作成される。両者の分析結果を比較すると、階層の上位にある内容、すなわちテキストの中心的内容とされる部分が一致しなかった。Meyer の分析法は複雑で時間がかかり、読後に時間をおいた後の想起や不正確な想起プロトコルの分析に適しているのに対して、Kintsch の分析法は読後直後の想起プロトコルの得点化に適しているとしている (Meyer, 1985a)。このように、意味構造の分析法はさまざまであり、異なる研究課題や研究者の問題意識に即したものが使用されている。

また、認知心理学における日本語文章を例とした構造の分析法に関して茂呂 (茂呂, 1982) が報告している。これは、英文テキストを対象として発達してきたさまざまな分析手法を概観して整理し、それらの手法に基づいて、テキストの伝達内容を命題単位に分解し、命題を基礎にしてテキストの統括性を構造として表現する手続きを論じている。

それに対し、日本の国語学領域には、日本語の文章の成立ちや構造を扱う文章論という独自の研究の伝統を持つ分野がある。ここでは、文脈依存性が高い (寺村ら, 1990)、主要部分が文の後ろにある (木下, 1981)、トピックセンテンスがない段落がある (佐久間, 1987) などという日本語の特殊性を十分に認識した上で、日本語の文章の構造が分析されている。

例えば、永野賢 (永野, 1986) は、文や段落の接続、首語の連鎖、陳述の連鎖、主要語句の連鎖、統括という観点に基づいて、文章の構造を分析している。寺村ら (寺村, 1990) も同様に、接続表現、指示表現、反復と省略、提題表現、陳述表現という観点から文章構造を分析している。接続表現の種類に若干の相違があるが、両者は基本的に同様の方式であり、文と文との関係と語の繰り返しなどの局所的な構造をつみあげて文章全体をひとまとまりのものとして捉えるものであり、本稿の用語では「意味構造」に相当する。また、これらの方法に基づいて、日本語の自然言語処理領域ではテキストの結束構造分析の研究もなされている。

他方、心理学における文章理解の認知的科学研究がさらにすすむと「文章理解の過程」と「文章理解に必要な知識」に関心が持たれるようになった (Britton *et al.*, 1985)。van Dijk は、認知心理学的視点をテキスト言語学に取り入れ、テキスト構造を生成や理解の過程から捉える「過程志向モデル」を提唱した。これは、文章を執筆する場合などのテキスト生成においては、初めに書き

手が持っている主要なアイディアがあり、それが段階的に細かい意味に展開されていくことによってテキストが生成される。逆に、テキストを読む場合は、提示されたテキストを読みながら、読み手がその内容を順次、消去・一般化・構成などのマクロ操作によって切り詰めていった局所的な意味を把握し、その局所的な意味をさらにまとめていくという過程を経てテキスト全体の意味である「マクロ構造」が得られるというモデルである (de Beaugrande *et al.*, 1981)。

しかし、実際は、このテキスト自身の「マクロ構造」は、その上部で、テキスト構造に関する知識に基づく「マクロ構造」に結びつき、テキストの理解や要約にはこの二つのマクロ構造が関わっているとしている。なお、van Dijk は、1980年の著書 (van Dijk, 1980) では、後者を特に「上部構造 (superstructures)」と呼んでいる。これは、テキストの内容を機能面から捉えた構造で、テキストの種類毎に共通な構造についてのスキーマである。それに対し、前者の「マクロ構造 (macrostructures)」は、個々のテキストの全体的な意味や概要を表わすものであり、両者が明確に区別されるようになった。van Dijk は前者の意味での「マクロ構造」という語を、構造というよりはむしろ、直観的なテキスト全体の要約あるいは中心的な意味を表わすものとして用いている (van Dijk, 1980)。しかし、この「マクロ構造」という語は、初期には van Dijk 自身も明確な区別をしていなかったように、テキスト全体の成立ち、あるいは機能面からみた「上部構造」に相当する意味で使用されている場合もある。このように用語の使用が統一されていないことがテキスト構造の研究の妨げの一つとなっていると考えられる。

2. 機能構造の分析

G. Crookes (Crookes, 1986) は、機能構造の妥当性のある分析法について検討し、実際に学術論文の「緒言」の機能構造を分析している。分析を妥当なものとするには、①分析の単位、②分析法、③分析対象となる資料の明記が必要であるとしている。

具体的には、①「文」を単位とし、②分析用のカテゴリとテキストへのカテゴリ付与の基準を明らかにし、カテゴリ付与の例を示すことが重要であるとしている。分析用カテゴリとして「重要性の確立」「既存研究の要約」「既存研究の不完全さの指摘」「論文の目的と範囲」を設定し、テキスト中の各文にそれぞれカテゴリを付与し、

テキスト毎にそのカテゴリの出現順を調べている。さらに、Crookes は、分析の客観性を確立するために2名の分析者の同一テキストに対する分析結果が一致した部分だけを用いて、構造を記述している。

また、機能構造の分析における分析カテゴリは研究者の問題意識に応じてさまざまな詳しきや観点から設定することができる。したがって、カテゴリ付与の妥当性だけでなく、分析用カテゴリ自体の妥当性も検討しなければならない。カテゴリとカテゴリ付与の双方の妥当性を示す方法として、複数分析者の同一テキストに対する分析結果の一致度を示している研究がみられた。

学術的なテキストの構造については、このほか、レトリック、科学社会学、科学史、研究過程論などでも研究されている。しかし、たとえば、科学史や科学社会学では個人の研究者の思想を分析するために構造を分析する (Kircz, 1991) など、それぞれ問題意識や観点が異なり、図書館・情報学における独自の観点から構造を捉える必要がある。たとえば、Farrow (Farrow, 1991) が主題分析は通常の記事理解とは異なる過程としたモデルを提示したように、関連領域の手法を援用しても、独自の観점에立って情報メディアの構造を捉える必要があるであろう。

C. 情報メディアの利用からみた特性と今後の展望

ここでは、図書館・情報学での情報メディアの構造分析のあり方と応用を展望する一つの方向として、情報メディアの利用に着目し、諸領域での研究成果を踏まえ、その特性を検討する。

図書館・情報学で主に対象とする学術情報メディアは、その特性の一つとして、基本的に同じ専門領域内での伝達を想定したメディアであり、長い間に形成された内容構成に関する慣習があることが挙げられる。特に学術論文は、その歴史のごく初期から (Holmes, 1991)、事実の記録と批判的議論という二つの性格を兼ね備えた内容構成をもち (Huth, 1982)、事実と議論の比率は時代と共に変化している (Bazerman, 1984)。また、特定研究者の原稿の詳細な検討から、社会の中にメディアの種類や性質に適した構成や論旨の展開の仕方についての暗黙の了解があり、編集過程はそれに従った内容構成を強める方向で機能していることも示されている (Myers, 1990)。

また、学術情報メディアの2番目の特性として、部分だけを読む、目的に応じて必要な部分だけを読む、ある

いは記載順とは異なる順序で読むといった利用がなされていることが挙げられるだろう。たとえば、公共図書館でのインタビュー (Prabha *et al.*, 1988)、心理学と人間工学分野の研究者の調査 (Dillon *et al.*, 1988)、物理学研究者へのインタビューと観察 (Bazerman, 1985)、電子工学研究者と電子工学技術者へのインタビュー調査 (Kirsch *et al.*, 1984) やシミュレーション作業 (Guthrie *et al.*, 1987) から、必要な部分だけを読んだり、目的や資料に応じてさまざまな読み方がなされていることが示唆されている。文献から情報を探し出す認知過程には通常の文章理解とは異なる問題解決モデルが有効である (Guthrie, 1988) とされ、学術論文の慣習的な構造についての知識は論文の内容の記憶を促進する (Samuels *et al.*, 1983) と報告されている。

さらに、3番目の特性として、学術情報メディアの利用においては、伝達内容の評価が重要であるということが挙げられる。専門的な情報の妥当性を真に評価できるのはその領域の専門家だけである (Kircz, 1991) といわれている。また、情報メディアによって伝えられている内容の妥当性や信頼性を評価しながら読むことは「批判的論評 (critical appraisal)」といい、特に臨床医学領域で関心が高まっている。臨床医学領域における批判的論評のガイドラインでは、論評の対象となっている論文で報告されている研究の種類に応じて、論文中で特定の役割を果たしている部分に着目するように指示されている (Sacket *et al.*, 1985)。また、論文の内的妥当性と外的妥当性を評価するためには、論文中の特定の要素間の整合性に着目することが示唆されている (Fletcher *et al.*, 1986)。

そのうえ、学術情報メディアでは、複数メディアの内容を利用者が比較、対照、統合して一つの解決を導き出すという利用がなされる。また、レビューの執筆や複数の調査結果を妥当性を吟味しながら統合するメタ分析 (Rosenthal, 1991) を行なうために複数のメディアの伝達内容を統合する形で利用されることもある。情報検索においても情報メディア間の関係づけも重要な観点になるだろう。

このように、学術情報メディアには利用という観点から以下のような特性がある。すなわち、①専門家間の情報の伝達のしかたとその構成に関する慣習がある、②部分だけの利用、③目的に応じた多様な利用、④内容の妥当性を評価しながら利用する、⑤複数のメディアの伝達内容を比較・対照・統合して利用する、などの特性があ

る。したがって、情報メディアの伝達内容がより有効に利用されることを促進するためには、これらの利用上の特性に適した方法で情報メディアの伝達内容の構造を分析し、その構造に適した情報システムや情報の提供方法を考えることが必要である。

すなわち、①構成に関する慣習に合致した方法で伝達内容を捉えて、提供することが望ましい。また、このような慣習に基づいて、伝達内容を分析するための一定の枠組みを設定することが可能であると考えられる。②部分利用と③目的に応じた多様な利用ということから、伝達内容の部分に着目したアプローチが適当であると考えられる。④妥当性の評価ということから、部分間の整合性や関係、部分と全体との関係に着目する必要があることがわかる。さらに、⑤複数メディアの利用、あるいは、情報検索という場を考慮すると、情報メディア間の関連付けを可能にし、一定の情報メディア間で共通の枠組みで伝達内容を分析することが必要であると考えられる。

以上より、情報メディアの伝達内容を構造として捉える、すなわち、伝達内容の部分、部分間の関係、部分と全体との関係に着目して情報メディアの伝達内容の構造を分析することは、このような利用上の特性に適したアプローチであると思われる。

また、情報メディアの伝達内容の構造を分析することを情報検索に応用する場合には、前提として構造の分析法の妥当性を確立することが必要である。作業負荷を考慮すると分析の自動化も必要であろう。検索結果を提示するには、上述のような目的に応じた多様な利用に対応する必要があり、他のメディアの内容との対比や統合も考慮したシステムとその基盤となるような構造分析の手法が必要であろう。さらに、伝達内容の妥当性の評価には、Fuller (Fuller, 1984) や Hibberd ら (Hibberd *et al.*, 1978) のように領域や研究方法を規定して機能面からみた詳細なカテゴリを設定して機能構造を分析することも一つの方法である。しかし、妥当性の評価にはなんらかの形で専門知識を取り込む必要もあると思われる。また、情報メディアの伝達内容の特性を捉えるために構造を分析するという観点からは、特に学術論文の規範的な構成といわれているものの源泉や経緯を検討することも必要であろう。

また、冒頭に述べたごとく、情報メディアの伝達内容には基本的特性として構造があり、その分析は情報メディアの伝達内容の特性を明らかにすると共に、その最も本質である「情報」にちかづくアプローチの一つと考え

られる。構造の分析は、情報メディア研究の基盤を提供し、さまざまな応用が考えられる。一方、情報メディアの伝達内容は人間の知的創造の成果を記録した非常に複雑なものであることから、その構造を分析する方法は一つに集約されるものではなく、さまざまな面から捉えられるべきであると考えられる。de Beaugrande ら (de Beaugrande *et al.*, 1981) がテキスト構造の分析について述べているのと同様に、分析の多様性が情報メディアの伝達内容の理解を深めることになるであろう。

ただし、これは無秩序な分析を奨励することではない。情報メディアの伝達内容の理解には伝達内容の構造の分析に関する研究の累積が必要である。それには、分析法に関して共通の用語を用い、分析法を詳細に記述すると共に、分析の立場や目的を明示し、同一の分析法による分析の累積を促進すると共に、他の分析法との位置づけを明確にすることが求められる。

本稿の執筆にあたり、ご指導くださいました慶應義塾大学文学部上田修一教授に感謝申し上げます。

- Allen, Bryce. Text structures and the user-intermediary interaction. *RQ*. Vol. 27, No. 4, p. 535-541 (1988)
- Allen, Bryce. Recall cues in known-item retrieval. *Journal of the American Society for Information Science*. Vol. 40, No. 4, p. 246-252 (1989a)
- Allen, Bryce. Propositional analysis: a tool for library and information science research. *Library and Information Science Research*. Vol. 11, No. 3, p. 235-246 (1989b)
- Allen, Bryce. The effects of academic background on statements of information need. *Library Quarterly*. Vol. 60, No. 2, p. 120-138 (1990a)
- Allen, Bryce. Knowledge organization in an information retrieval task. *Information Processing and Management*. Vol. 26, No. 4, p. 535-542 (1990b)
- Anderson, James D. Indexing systems: extensions of the mind's organizing power. *Information and Behavior*. Vol. 1, p. 287-323 (1985)
- Bazerman, Charles. Modern evolution of the experimental report in physics; Spectroscopic articles in *Physical Review*, 1893-1980. *Social Studies of Science*. Vol. 14, No. 2 p. 163-196 (1984)
- Bazerman, Charles. Physicists reading physics. *Written Communication*. Vol. 2, No. 1, p. 3-23 (1985)
- Beghtol, Clare. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*. Vol. 42, No. 2, p. 84-113 (1986)
- Belkin, N. J.; Robertson, S. E. Information science and the phenomenon of information. *Journal of the American Society for Information Science*. Vol. 27, No. 4, p. 197-204 (1976)
- Bonzi, S.; Liddy, E. The use of anaphoric resolution for document description in information retrieval. *Information Processing and Management*. Vol. 25, No. 4, p. 429-441 (1989)
- Bovair, Susan; Kieras, David E. "12. A guide to propositional analysis for research on technical prose". *Understanding Expository Text*. Lawrence Erlbaum, 1985, p. 315-362.
- Britton, Bruce K.; Black, John B. "1. Understanding expository text: from structure to process and world knowledge". *Understanding Expository Text*. Lawrence Erlbaum, 1985, p. 1-9.
- Buxton, A. B.; Meadows, A. J. Categorization of the information in experimental papers and their author abstracts. *Journal of Research Communication Studies*. Vol. 1, p. 161-182 (1978)
- Craven, Timothy C. Sentence dependency structures in abstracts. *Library and Information Science Research*. Vol. 10, No. 4, p. 401-411 (1988)
- Craven, Timothy C. Condensed representation of sentences in graphic displays of text structures. *Journal of Documentation*, Vol. 46, No. 4, p. 339-352 (1990)
- Cremmins, E. T. *The Art of Abstracting*. Institute for Scientific Information, 1982. 150 p.
- Crookes, G. Towards a validated analysis of scientific text structure. *Applied Linguistics*. Vol. 7, No. 1, p. 57-70 (1986)
- de Beaugrande, Robert; Dressler, Wolfgang. *Introduction to Text Linguistics*. London, Longman, 1981, 270 p. (Longman Linguistics Library; 26) (池上嘉彦: 三宮郁子訳. テキスト言語学入門. 東京, 紀伊国屋書店, 1984. 348 p.)
- Dea, W.; Belkin, N. J. Beyond the sentence: clause relations and textual analysis. *Informatics 3; Proceedings of a Conference held by the Aslib*. London: Aslib, 1978, p. 67-84.
- Dillon, Andrew; Richardson, John; McKnight, Cliff. Towards the development of a full-text, searchable database: implications from a study of journal usage. *British Journal of Academic Librarianship*, Vol. 3, No. 1, p. 37-48 (1988)
- Farradene, J. Relational indexing, part 1-2. *Journal of Information Science*, Vol. 1, No. 5, p. 267-276; No. 6, p. 313-324, (1980)
- Farrow, J. F. A cognitive process model of docu-

- ment indexing. *Journal of Documentation*. Vol. 47, No. 2, p. 149-166 (1991)
- Fidel, Raya. Writing abstracts for free-text searching. *Journal of Documentation*. Vol. 42, No. 1, p. 11-21 (1986)
- Fletcher, R. H.; Fletcher, S. W.; Wagner, E. H. 臨床のための疫学. 久道茂ほか訳. 東京, 医学書院, 1986.
- Fredriksen, C. H. Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology*, Vol. 7, p. 371-458 (1975)
- Frohmann, Bernd. Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*. Vol. 46, No. 2, p. 81-101 (1990)
- Fuller, Sherrilynne Shirley. Schema theory in the representation and analysis of text. Univ. Southern California, 1984. 189 p. PH. D thesis, available from U.M.I. order No. DA8500206.
- Guthrie, John T.; Kirsch, Irwin S. Distinctions between reading comprehension and locating information in text. *Journal of Educational Psychology*. Vol. 79, No. 3, p. 220-227 (1987)
- Guthrie, John T. Locating information in documents: examination of a cognitive model. *Reading Research Quarterly*. Vol. 23, No. 2, p. 178-199 (1988)
- Hahn, Udo. Topic parsing: accounting for text macro structures in full-text analysis. *Information Processing and Management*. Vol. 26, No. 1, p. 135-170 (1990)
- Hibberd P. L.; Meadows, A. J. Information contained in clinical trial reports. *Journal of Information Science*. Vol. 2, No. 3/4, p. 165-168 (1980)
- Hoffmann, Eliahu. Defining information: an analysis of the information content of documents. *Information Processing and Management*. Vol. 16, No. 6, p. 291-304 (1980)
- Holmes, Frederic L. "6.: Argument and narrative in scientific writing", *The Literary Structure of Scientific Argument: Historical Studies*. Univ. of Pennsylvania Press, 1991, p. 164-181.
- Huth, Edward J. How to Write and Publish in the Medical Science. ISI Press, 1982. 203 p.
- International Organization for Standardization/TC 97, prep. *Information Processing: Text and Office Systems: Standard Generalized Markup Language (SGML)*, 1st ed. ISO, 1986, ISO 8879-1986.
- Janos, Jiri. Theory of functional sentence perspective and its application for the purposes of automatic extracting. *Information Processing and Management*, Vol. 15, No. 1, p. 19-25 (1979)
- Jones, Kevin P. How do we index?: a report of some ASLIB Informatics Group Activity. *Journal of Documentation*. Vol. 39, No. 1, p. 1-23 (1983)
- 神門典子. 構成要素カテゴリーを用いた情報メディア内部構造分析の試み: 原著論文を例として. 平成2年度慶應義塾大学文学研究科修士論文. 1991, 227 p.
(結果の概要は次を参照. 神門典子. 構成要素カテゴリーを用いた原著論文の内部構造分析. 情報処理学会研究報告, Vol. 92, No. 32, p. 39-46 (1992) (92-FI-25))
- 木下是雄. 理科系の作文技術. 中央公論社, 1981. 244 p.
- Kintsch, Walter; van Dijk, Teun A. Toward a model of text comprehension and production. *Psychological Review*. Vol. 85, No. 5, p. 363-394 (1978)
- Kircz, Joost G. Rhetorical structure of scientific articles: the case for argumentational analysis in information retrieval. *Journal of Documentation*, Vol. 47, No. 4, p. 354-372 (1991)
- Kirsch, I.; Guthrie, J. Adult reading practices for work and leisure. *Adult Education Quarterly*. Vol. 34, No. 4, p. 213-232 (1984)
- Lancaster, F. W. *Indexing and Abstracting in Theory and Practice*. London, Library Asson., 1991, 328 p.
- Liddy, E.; Bonzi, S.; Katzer, J.; Oddy, E. A study of discourse anaphora in scientific abstracts. *Journal of the American Society for Information Science*. Vol. 38, p. 255-261 (1987)
- Liddy, Elizabeth DuRoss. The discourse-level structure of empirical abstracts: an exploratory study. *Information Processing and Management*. Vol. 27, No. 1, p. 55-81 (1991)
- Maeda, Takashi, An approach toward functional text structure analysis of scientific and technical documents. *Information Processing and Management* Vol. 17, No. 6, p. 329-339 (1981)
- Meyer, Bonnie J. F. "2: Prose analysis: purposes, procedures, and problems". *Understanding Expository Text*. Lawrence Erlbaum, 1985a, p. 11-63.
- Meyer, Bonnie J. F. "10; Prose analysis: purposes, procedures, and problems; pt. 2". *Understanding Expository Text*. Lawrence Erlbaum, 1985b, p. 269-304.
- Milas-Bracovic, Milica; Zajec, Jasenka. Author abstracts of research articles published in scholarly journals in Croatia (Yugoslavia): an evaluation. *Libri*. Vol. 39, No. 4, p. 303-318 (1989)
- 茂呂雄二. 日本語文章の形式的表示 (2); 統括性を中心に. *読書科学*. Vol. 26, No. 1, p. 24-35 (1982)

- 武者小路澄子. 原著論文と抄録の関係における質的分析: 情報の圧縮化へのアプローチ. *Library and Information Science*. No. 26, p. 1-29 (1988)
- 武者小路澄子. 情報メディアの形成過程: 著者抄録を中心とした質的分析. *Library and Information Science*. No. 27, p. 15-36 (1989)
- Myers, Greg. *Writing Biology: Texts in the Social Construction of Scientific Knowledge*. Madison, Univ. of Wisconsin Press, 1990.
- 永野賢. 文章論総説: 文法論の考察. 東京, 朝倉書店, 1986. 379 p.
- O'Connor, John. Answer-passages retrieval by text searching. *Journal of the American Society for Information Science*. Vol. 31, No. 3, p. 227-239 (1981)
- Oddy, Robert N.; Liddy, Elizabeth DuRoss; Balakrishnan Bhaskaran; Bishop, Ann; Elewononi, Joseph; Martin, Eileen. Towards the use of situational information in information retrieval. *Journal of Documentation*. Vol. 48, No. 2, p. 123-171 (1992)
- 大橋保夫編. クロード・レヴィ=ストロース日本講演集 構造・神話・労働. 東京, みすず書房, 1979, 188 p.
- 大沼美佐. 索引作業の認知科学的分析. 昭和 63 年度慶應義塾大学文学部図書館・情報学科卒業論文, 1988.
- Paice, Chris D. Constructing literature abstracts by computer: techniques and prospects. *Information Processing and Management*. Vol. 26, No. 1, p. 171-186 (1990)
- Prabha, Chandra; Bunge, John; Rice, Duane. Aspects of nonfiction book use: a study of public library patrons. *Library and Information Science Research*, Vol. 10, No. 2, p. 177-193 (1988)
- Rosental, Robert. Meta-analysis: a review. *Psychosomatic Medicine*, Vol. 53, p. 247-271 (1991)
- Rowley, Jennifer E. *Abstracting and Indexing*, 2nd ed. London, Clive Bingley, 1988. 181 p.
- 佐久間まゆみ. 「文段」認定の一基準 (I) 提題表現の統括. *文芸言語研究言語篇*, No. 11, p. 89-135 (1987)
- Sackett, David L.; Haynes, R. Brian; Tugwell, Peter. *Clinical Epidemiology: a Basic Science for Clinical Medicine*. Boston, Little Brown, 1985. 370 p.
- Samuels, S. Jay; Tennyson, Robert; Sax, Lynn; Mulcahy, Patricia; Schermer, Nancy; Hajovy, Halyna. Adults' use of text structure in the recall of a scientific journal article. *Journal of Educational Research*. Vol. 81, No. 3, p. 171-174 (1988)
- SIST-01 科学技術庁振興局管理課情報室編. 科学技術情報流通技術基準 抄録作成. 東京, 日本科学技術情報センター, 1980. SIST-01-1980, 11 p.
- 杉原寛子. 抄録作成者の情報処理: 抄録作成におけるマクロロールの役割. *Library and Information Science*, No. 23, p. 63-75 (1985)
- 寺村秀夫; 佐久間まゆみ; 杉戸清樹; 半澤幹一. ケーススタディ日本語の文章・談話. 東京, 桜楓社, 1990. 189 p.
- Tibbo, Helen R. Abstracting across the disciplines: a content analysis of abstracts from the natural sciences, the social sciences, and the humanities with implications for abstracting standards and online information retrieval. *Library and Information Science Research*, Vol. 14, p. 31-56 (1992)
- 内田伸子. “5.3. 文章理解と知識”. *認知心理学講座 3: 推論と理解*. p. 158-179 (1982)
- van Dijk, Teun A. *Macrostructures: an Interdisciplinary Study of Global Structures in Discourse, Interaction, and Cognition*. Hillsdale, Lawrence Erlbaum Assoc., 1980, 317 p.
- Weiner, J.M. Text analysis and basic concept structure. *Information Processing and Management*. Vol. 19, No. 5, p. 313-319 (1983)
- Wilson, Patrick. Some fundamental concepts of information retrieval. *Drexel Library Quarterly*. Vol. 14, No. 3, p. 10-24 (1978)