日本語文構造解析による自動インデクシング方式†

絹川博之†† 木村睦子†††

事実検索を指向した情報検索システムの運用においては，(1)情報入力蓄積作業の省力化・標準化と(2)検索精度の向上をどのように実現するかが問題である．扱う情報が，漢字かな混り日本語文である場合は，漢字入力の困難さから，この問題は，より大きなものとなる．本論文では，この問題を解決することを目的に，記事文検索を例に，検索用インデックスであるキーワードの付与を現在の人手付与から効率の高い自動化方式に改善し，かつ検索精度向上のため，情報の意味内容表示子として，各キーワードにロール（記事文中における主体，客体等の意味的役割）を自動付与（自動インデクシング処理）する新しい方式を提案し，評価している．すなわち，本研究では(A)名詞と述語の依存関係に着目し，かつ意味情報を利用した日本語文構造解析法と，(B)文構造を述語を軸に表形式にまとめた文型表を参照する方法によるプログラム構造を提案し，自動処理方式を実現している．また(C)漢字プリンタ，漢字ビデオ端末を用いた会話型校正機能をも具備した具体的システムを開発している．この結果，従来の人手によるキーワードの付与に比べ，(i)インデックス付与作業の所要時間は1/3以内，(ii)検索精度は，検索モレを増加させることなく，検索誤り発生率を20～30%低減（改善），(iii)情報蓄積のターン・アラウンド時間は，1/2以内となり，上記の目的を実現することができた．

1. まえがき

事実検索や，それに近い使い方をする情報検索システムの運用においては，(1)情報入力蓄積作業の省力化・標準化と，(2)検索精度の向上をどのように実現するかが問題である．すなわち，事実検索指向の情報検索においては，検索用インデックスとして，単にキーワードを用いるだけでは，検索結果に誤りが混り，検索精度の水準を確保することは，難しい．これに対処するためには，入力情報に，その情報の意味内容表示子をインデックスとして付与し，検索論理式に組み込めるようにすることが考えられる．ところが，キーワードに加えて，意味内容表示子までを人手で付与して，情報蓄積作業を行うことは，大量の情報の蓄積を前提とするシステムでは，実用的でない．また，扱う情報が，漢字かな混り日本語文である場合は，漢字入力の困難さから，情報蓄積の際のデータ量を少なくすることが，必要とされている．

このような状況の中で，英数字，カナ表記情報においては，インデックスのうちのキーワードの自動付与が実用化されてきているが[7]，漢字を含む日本語文情報においては，文中の言葉の出現傾向を利用[1),2)]したキーワード付与が，実験されているにすぎない．また，いずれの場合も，問題点(1)の解決を主眼にしている．これに対して，本論文では，問題点(1)，(2)の両方を解決することを目的に，記事文検索を具体例として進め，検索用インデックスであるキーワードの付与を現在の人手付与から効率の高い自動化方式に改善し，かつ，検索精度向上のため，情報の意味内容表示子として，各キーワードにロール(記事文中における主体，客体等の意味的役割）を自動付与する新しい方式を提案し，評価している．すなわち，本研究では，(A)名詞と述語の依存関係に着目し，かつ，意味情報を利用した日本語文構造解析法，(B)文構造を述語を軸に表形式にまとめた文型表を参照する方法によるプログラム構造，を提案し，自動処理方式を実現している．また，(C)漢字プリンタ，漢字ビデオ端末を用いた会話型校正機能を具備した具体的システムを開発し，評価している．

以下，本論文では，本自動インデクシングの設計目標と方針(第2章)，設計の基本概念として定式化した日本語文構造解析方式(第3章)，本システム構成とその評価（第4章）について述べる．

2. 設計目標と方針

検索精度の確保・向上のため，情報の意味内容表示子を使用する方式を提案したが，政治・経済・外交関連の記事文検索においては，例えば，ある人物が起した事件とか，同一人物に対する評価とかに関する記事を区別することの必要性から，ロール付きのキーワードをインデックスとして，使用する方式を採用した．

---

† Automatic Indexing System Utilizing Japanese Sentence Analysis HIROSHI KINUKAWA (Systems Development Laboratory, Hitachi, Ltd.) and MUTSUKO KIMURA (Institute of Behavioral Sciences).

†† (株)日立製作所システム開発研究所

††† (財)計量計画研究所

ロールは，キーワードが，それを含む文中において，①（有意志の）主体，②（有意志の）客体，③時，④場所，⑤活動，⑥①～⑤以外の主題，の6種のいずれであるかを識別するものでのる．本自動インデクシング方式は，キーワードとロールの自動付与を行うものであり，その設計目標としては，第1に各種テーブルのデータの補追によりシステムの性能向上が容易にできる方式を開発することである．本自動インデクシング方式では，自然語文解析を基本としている．従来の多くの自然語文解析で行われているように，言語データと解析手続を分離する方式を基礎とすることにより，解析方式の保守性を向上させようとするものである．また，言語データの保守には，出現頻度を考慮することとし，その頻度が少なく，テーブル類への補追が，処理誤りを引き起すものについては，補追せず，人間の援助を得て処理する方式をとった．

第2の目標は，漢字かな混り日本語文の情報検索におけるインデクシング方式として，記事文検索以外にも汎用的に使用可能な方式を開発することである．今後，日本語文情報の検索のニーズは，広い分野において高まり，システムの運用に当っては，人手作業を僅少とすることが必須であると考えるからである．また，自動インデクシング・システムは，情報検索システムの情報蓄積プロセスを構成する要素として，他の要素と共存できることが必要である．

第1の目標に対しては，日本語の文構造が，名詞と述語の依存関係[3),4)]（係り受け関係）により規定されることと，その関係における格助詞の果す機能に着目して，C. J. Fillmore[5)]の格文法の考え方を基礎に，ロールを含めた日本語文構造モデルを提案した．また，このモデルに基づき，文型表を作成して，解析手続と解析に用いる言語データを分離する方式を提案した．

第2の目標であるインデクシング・システムの汎用性に対しては，日本語の特性に着目して，用語辞書を対象分野に依存しない付属語表と，対象分野に依存する自立語辞書に分離するとともに，キーワード抽出手続とロール付与手続を分離する方式を提案した．また，自動処理結果の出力形式は，入力した原データに処理結果を付加する形式を採用した．

3. 日本語文構造解析方式

日本語文構造と意味構造の関係から，ロールを付与する新しい方式を提案した．以下，この詳細を述べる．

3.1 日本語文構造とロール

日本語文における単文の表層構造の基本的骨格は，

〈名詞文節〉$_1$…〈名詞文節〉$_n$〈述語〉　　(3.1)

となっている[6)]．実際の文では，〈名詞文節〉に連体修飾語が係っていたり，名詞の並列関係が存在しうるが，基本直骨格を考えているので，除いている．ここで，nは，正整数で述語に直接従属する文節数を表わす．日本語の特性は，修辞的な表現を除くと，次のようになる．

（1） 述語が，文末に来る．

（2） 述語以外の名詞文節には，次の特性がある．

（a） 名詞文節は，「〈名詞〉〈格助詞〉」という構成を有する．ここで，格助詞（主として，ガ，ニ，ヲ）に代るものとして，係助詞を多く用いられる．

（b） 語順が，不定である．

（c） 語の省略が可能である．

C. J. Fillmore[5)]の格文法の考え方により，自然語文表現の意図する意味構造（深層格構造）は，述語の格関係として，$\boldsymbol{P}_J(e_1, e_2, \cdots, e_{n_J})$と表わすことができる．ここで，$\boldsymbol{P}_J$は，各述語の格関係の識別子であり，動詞形容詞等の述語Jの代表的表現である．また，e_i $(1 \leqq i \leqq n_J)$は，名詞の格（例えば，[主格]，[対象格]，[場所格]）である．

文の表層構造から深層格構造への対応関係が，明確になれば，文の言わんとする意味が，明確となる．

これを，定式化すると

[文の意味]＝[文の深層格構造]
＝$\boldsymbol{M}$([文の表層構造])　　(3.2)

となる．ここで，$\boldsymbol{M}$は，マッピング函数である．

本システムのロールは，キーワードの文中における意味的役割であり，式(3.2)の定式化に対応させると，マッピング函数 $\boldsymbol{M}$ に相当する．すなわち，

$\boldsymbol{M}$＝[①主体，②客体，③時，④場所，⑤活動，⑥主題]　　(3.3)

となる．式(3.3)のように定式化すると，深層格構造の［格関係］，各［格］は，$\boldsymbol{M}$の射影となる．

$\boldsymbol{M}$/⑤活動([文の表層構造])＝[格関係]　(3.4)

$\boldsymbol{M}$/①主体([文の表層構造])＝[主格]　(3.5)

$\boldsymbol{M}$/④場所([文の表層構造])＝[場所格]　(3.6)

etc. である．ここで $\boldsymbol{M}$/⑤活動は，$\boldsymbol{M}$の「⑤活動」による射影を表わす．

深層格構造に，[格関係]とその[格関係]を構成する要素というレベルの違いがあり，その[格関係]を

規定するのは，〈述語〉であることから式(3.4)～式(3.6)は，次のように変形される．

M/⑤活動 (〈述語〉)=[格関係] (3.4)′

M/①主体 (〈名詞文節〉$_1$…〈名詞文節〉$_n$*/〈述語〉)
=[主格] (3.5)′

M/④場所 (〈名詞文節〉$_1$…〈名詞文節〉$_n$*/〈述語〉)
=[場所格] (3.6)′

ここで，*印は，〈述語〉が制限条件として，働いていることを示す．

式 (3.4)′～式 (3.6)′ において

〈名詞文節〉$_1$…〈名詞文節〉$_n$

の文節列のうちの一つの〈名詞文節〉$_i$ が，実際には，[主格] とか [場所格] とかの，ある格に対応する．

以上の日本語文構造モデルを基礎に，ロールを規定する要件を分析した．その内容は，次の通りである．

(1) ⑤活動：動作を表わす動詞に限られ，〈述語〉の部分集合で，個々の述語を見れば，動作を表わすか否かは定まる．

(2) ③時：他のロールとの多義性は，あり得ない．

(3) ①主体，②客体，④場所，⑥主題については，次のごとくである．

(a) 有意志体は，ロール①かロール②となる．

(b) 場所は，ロール④にしかなり得ない．

(c) 無意志体は，ロール⑥にしかなり得ない．

(d) 国名などのように，組織体として意志を有したり，その国の領有する地名として場所を示したりする語がある．

(e) 航空機などのように，それを操作する人間を含め擬人化されたり，場所になったり，人間が作るものとして物質名になったりする語がある．

(f) (a)～(c)の区別は，名詞文節中の名詞の意味により定まる．

(g) (d)，(e)の多義性は，それがいかなる〈述語〉の支配下にあるかにより，有意志体か，場所か，その他かは，定まる．

(h) (a)のロール①か，ロール②かの多義性は，式 (3.5)′，式 (3.6)′ 等の定式化により

(i) いかなる〈述語〉の支配下にあるか

(ii) (i)の支配下で，名詞文節にいかなる格助詞が付いているか (**図1**例(1)，例(2)参照)

の2条件で定まる．ここで，日本語の特性として格助詞が，表層構造上の主語，目的語等を定める．条件(ii)は，述語との関係から，深層格構造上の格が，表層構造上では，主語や，目的語として，具現するということによるものである．

(1) 米国が イギリスを支配する。
主体 客体 活動

(2) 米国が 地中海を支配する。
主体 場所 活動

(3) 米国が 石油を 支配する。
主体 主題 活動

(4) 排外思想が 米国を支配する。
主題 客体 活動

図1 ロール付与の例

Fig. 1 Examples of role-setting to key words.

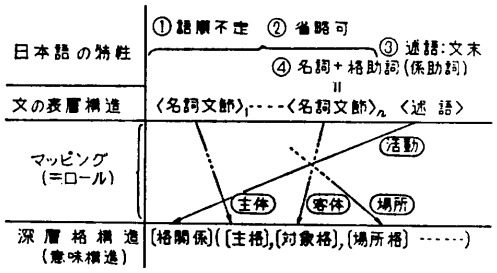

図2 日本語文の基本構造とロール

Fig. 2 Relationship between surface case structure and case structure in Japanese sentence.

(4) 以上より，ロールを規定する要件は，

(a) 名詞文節を支配する述語

(b) 名詞文節構成名詞の意味

(c) 名詞文節に付接する格助詞

の3つである．以上の関係を**図2**に示す．

3.2 文構造解析の基本的な考え方

ロールを自動付与するには，第3.1節に提案した日本語文構造モデルから，文における [格関係] と，各 [格] との依存関係 (支配従属関係) を明らかにすることが必要である．このため，本自動インデクシングにおける日本語文構造解析では，支配従属文法 (Dependency Grammar) に則り，文全体のレベルから，述語の性質とそれに従属する名詞を，Top-down 方式で決定していく方式を採用している．従来，句構造文法，変形文法，遷移網文法などに基づく多くの構文解析法では，個々の単語を組み合わせて，句または節を構成していく方法，すなわち，Bottom-up 方式を採用していたが，本方式では，ロール付与という目的に鑑み，Top-down 方式を採用している．これは，Bottom-up 方式で構文分析を行うと，述語に直接従属する語以外の枝葉の部分に，多くの多義性を生じ，ロール付与という目的にとっては，無用の手間と混乱をきたすことになるからである．

第3.1節で，名詞の意味分類が，必要であることを述べた．本自動インデクシングにおいては，記事文中に出現する言葉について，名詞と述語の依存関係を分析する解析手続が高い処理効率を保持しつつ，ロール付与できることを主眼に，次に7種に意味分類している．

(i)組織体，(ii)人名，(iii)資料名，(iv)地名，(v)動作，(vi)物品名，抽象概念など((i)～(v)，(vii)以外)，(vii)時

第3.1節および以上の考え方に基づき，ロールの自動付与処理を行うために，述語類を次の点から分類した．すなわち，述語が支配する名詞格情報（名詞の意味分類と格助詞と組にした情報）と，付与すべきロールの違いにより，述語類を分類した．(**図3**) この分類に基づき文型表 (**図4**) を作成した．本システムでは，記事文中に出現する述語約 5600 語を 586 に分類 (この分類を格支配情報(A4情報)と呼んでいる.) した．各分類に属する述語は，通常，複数の文型を取り得るものであり，本文型表には，合計 1686 文型が定められている．また，1つの文型は，最大4個の文型要素（当該分類に属する述語が，支配する名詞格情報と，付与すべきロールとを対にしたもの）から構成される．例えば，「支配する」とか「攻撃する」とかが属する第 46 分類の述語は，6個の文型をとり得，いずれの文型も，2個の文型要素からなっている．図4の第 46 分類の1番目の文型は，(ガ，A，①) と，(ヲ，1，②) という文型要素からなることを示している．この三ツ組の第1のコードは，格助詞，第2のコードは，意味分類，第3のコードは，付与すべきロールを示し，意味分類の「A」は，「組織体」，「人名」のいずれでもよいことを示す．

本自動インデクシング処理では，この文型表を参照することにより，文構造解析を行い，ロールを付与する手順を実現した．

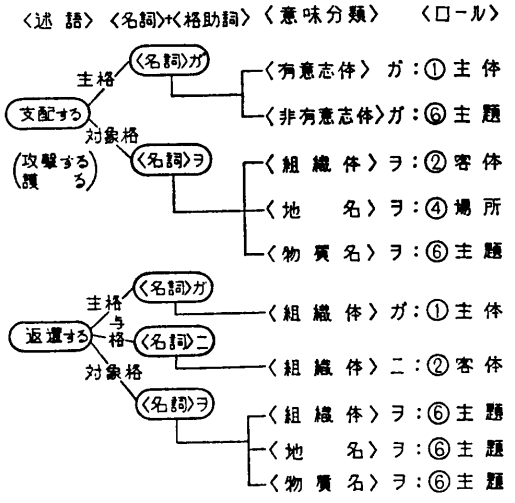

図3 述語とロールとの関係

Fig. 3 Relationship between predicate verb and roles.

A4情報	文型要素 名詞格情報 格	B	ロール	文型要素 名詞格情報 格	B	ロール	文型要素 名詞格情報 格	B	ロール	文型要素 名詞格情報 格	B	ロール
1	ガ	1	①									
⋮												
46 (支配する 攻撃する 護る etc.)	ガ	A	①	ヲ	1	②						
	ガ	A	①	ヲ	4	④						
	ガ	A	①	ヲ	6	⑥						
	ガ	6	⑥	ヲ	1	②						
	ガ	6	⑥	ヲ	4	④						
	ガ	6	⑥	ヲ	6	⑥						
⋮												
586												

(注) Bコード：意味分類をいい，その値は次の通り
1：組織体，2：人名，3：資料名，4：地名
5：動作，6：1～5,7以外 7：時，A：1または2（有意志体）

図4 提案した文型表

Fig. 4 Proposed sentence pattern table.

4. システムの構成とその評価

第3章で示した考え方に基づき，**図5**に示す構成の自動インデクシング・システムを試作した．その主たるハードウェアを**表1**に示す．

4.1 プログラム・システム構成

本システムは，次の3つの処理ステップからなっており，その機能を以下に述べる．

(1) 文節構成語の認定
(2) 日本語文構造解析によるロール付与
(3) 会話型校正サブシステム

これらを構成するプログラム性能を**表2**に示す．

4.1.1 文節構成語の認定[10] (図6参照)

(1) '文節' 分割[9]：平がなから非平がなに変わる

表1 主たるハードウェアの概要

Table 1 Hardware configuration.

項番	装置名称	仕様概要と用途
1	中央処理装置	主メモリ：384 KB，システム・ミックス 3.6 μs
2	磁気ディスク	平均アクセス時間：72.5 msec，シソーラス，文型表の蓄積媒体
3	漢字プリンタ	印字速度 700 行／分，処理結果および校正リスト作成用
4	漢字ビデオ端末	表示画面 40 字／行×12 行，会話型校正用

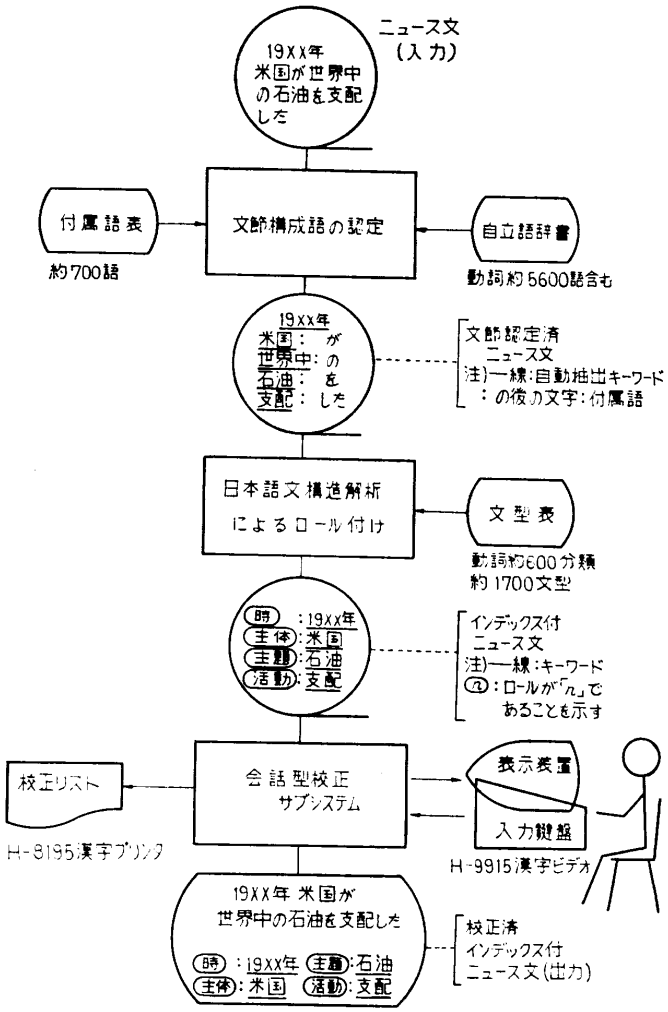

図 5 自動インデクシング・システムの構成
Fig. 5 Configuration of the automatic indexing system.

表 2 システム構成プログラムの容量と性能
Table 2 Size and performance of the programs.

項番	処理プログラム	ステップ数	最大所要メモリ	性能
1	文節構成語の認定	KS 3	KB 60	240 msec/文節
2	文構造解析によるロール付与	12	120	650 msec/センテンス
3	辞書・テーブル類保守プログラム群	6	33	—
4	会話型校正サブシステム	6	132	—
5	編集出力，データ収集用ユーティリティ	11	84	—
6	合計	38	132	—

点および記号類の所で，入力文字列を分割する．

(2) 自立語認定（含：キーワード抽出）

(a) '文節' 構成文字列を，自立語辞書の見出しと最長優先一致照合させ，一致文字列の表わす語を認定自立語とする．

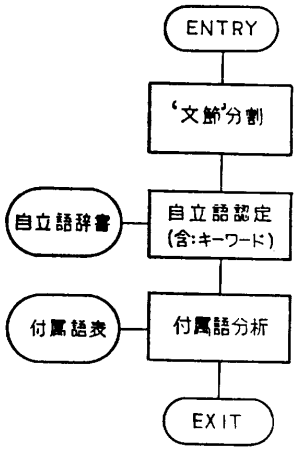

図 6 文節構成語の認定手順
Fig. 6 Segment recognition process.

(b) 自立語辞書には，名詞類約 13 万語，動詞・形容詞等の述語約 5600 語，その他文構成上必要な言葉を収録した．

(c) 辞書収録語には，次の解析用情報を与えた．

(i) A1情報：体言類は3種類，用言類は4種類，その他は3種類に品詞分類するための類別情報．

(ii) A2情報：自立語と付属語，用言の語幹と語尾，付属語同志の連接の仕方を 75 種類に分類するための種別情報．

(iii) A3情報：接頭，接尾辞の識別情報．

(iv) A4情報：述語の格支配情報であり，586 分類のいずれであるかの識別情報．(第 3.2 節参照)

(v) Bコード：名詞の意味分類コード．(第 3.2 節参照)

(3) 付属語表を使用した付属語分析[8),10)]

付属語表を使用することにより，付属語分析を行い，次のC1，C2，C3およびDの各情報を認定する．

(a) C1情報：名詞の格（格助詞の分類）

(b) C2情報：述語の態；能動，受身，使役

(c) C3情報：下に示す文節末形態の区別

(i) 体言文節：述語への直接従属，体言に対する修飾，または並列関係の区別．

(ii) 用言文節：活用形の区別．連用中止，連体，引用 etc.

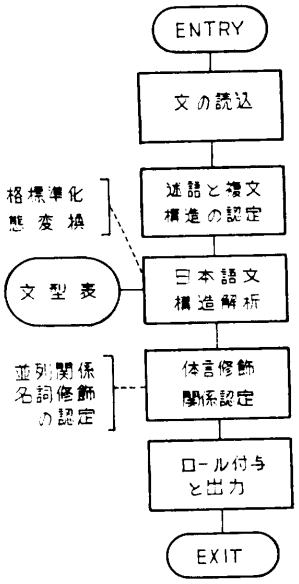

図 7 文構造解析によるロール付与手順

Fig. 7 Role-setting process utilizing Japanese Sentence Analysis.

（d） D情報：付属語が付接して初めて品詞の定まる言葉に対する処理情報.

(例) サ変語幹名詞は，「スル」の活用形が付接して初めて，動詞となる.

「説明」+「する」⟶「説明する」

4.1.2 文構造解析によるロール付与[10]（図7参照）

（1） 1文中の述語をA1情報により認定し，C3情報で，複文構造の認定を行う．記事文検索の場合は，引用関係と形式名詞関係の2つの複文構造のみ，それを示すリンクを与えることにし，他は，単なる列記とした．ここで，形式名詞関係とは，次の例の下線部のような場合を指す.

(例) 政府は，対米輸出を規制することを決定した.

（2） (1)の述語のA4情報により，文型表から該当文型を取得し，述語に近い位置の文節から文頭へという順に，名詞文節の述語に対する従属関係を，BコードとC1情報を用いて分析する方法を採用した.

この分析に先立って，次のことを行う.

（a） 使役態：使役態用の文型の探索を準備する.

（b） 受身態：名詞文節の格助詞を能動態のそれへと変換する.

（c） 係助詞の格助詞への変換

（d） 連体形用言：A4情報で得た文型から，「ガ」「ヲ」，「ニ」格のいずれかの文型要素1つを除いた連体形用言照合用の文型を作成する.

（3） 体言の修飾関係認定：連体形用言，並列関係，名詞間修飾関係については，それらの直後の名詞文節に係るものとした．これは，ロール付与の原則を，記事文検索において『体言を修飾する文節のロールは，述語に直接従属する体言のロールと同じとする.』としたことによる.

（4） ロール付与と出力：(1)～(3)の結果により，ロールを付与する．支配従属関係の定まらない名詞文節は，名詞のBコードとC1情報により，ロールを定めることとした．辞書未収録等で，情報の定まらないものは，ロール付与不能扱いとした.

4.1.3 会話型校正サブシステム

自動処理結果の確認と修正のためのサブシステムである．自動処理結果を，漢字プリンタで校正リストとして作成し，校正者がそれに修正情報を記入し，校正オペレータが修正済み校正リストを見ながら，漢字ビデオ端末を用いて，会話的にファイルを修正できるシステムである.

4.2 処理性能

（1） 記事文検索における蓄積用記事281件（1225センテンス）を対象として実験を行い，自動処理精度を評価した．精度は，次の2つの条件に依存している.

（a） 用語辞書の出現語彙に対するカバー状況

（b） 文型表の記事文構文に対するカバー状況

（2） 上記第(1)項の2つのカバー率を，ともに90% とした時の精度は，自動処理精度の分析から，次のように算定された.

（a） 文節構成語認定精度：85～90%，キーワード抽出精度は，本値と同じである.

（b） ロール付与精度：80～85%

$$\left(\text{処理精度}=\frac{\text{正処理数}}{\text{処理対象総数}}\times 100\right)$$

自動処理結果例を図8に示す.

（3） 本自動インデクシングの処理限界を次に示す.

（a） 辞書未収録語は，ロール付与不能となる．特に述語が未収録である時は，1単文全体について，文構造解析が不可能となる.

（b） 述語を複数含む文の複文構造の多義性の選択を本システムでは行っていない．記事文については，このことのロール付与精度に対する悪影響は，数%以下で，辞書や文型表の不備による影響と比べて，極めて小さい.

```
                                                    P. 35
   "自動インデックス・システム" 校正リスト '79年1月9日
   記事番号：1102827○ ア 文番号：1
<入力原文>
 760725ポルトガル政府，同国の元植民地東チモールのインドネ
 シア併合を正式に承認
<処理>   ID作成区分（ ）
LK 番号 R キーワード 番号 R キーワード 番号 R キーワード
主文   1 ⑤承認
       2 ③760725*1 3 ①ポルトガル政府
       4 ⑤同国⑥元⑤植民地⑥東⑥チモール⑥インドネシア⑥併合
       5 ⑥正式
<入力原文>
 760727フィリピン・ロムロ外相，沢木駐比大使と日比友好通商
 航海条約の処理について会談
<処理>   ID作成区分（ ）
LK 番号 R キーワード 番号 R キーワード 番号 R キーワード
主文   1 ①フィリピン ①ロムロ外相        2 ⑤会談
       3 ③760727   4 ⑥日⑥比⑥友好⑥通商航海条約
                     ⑥処理
シ無*2 5 ○沢木駐比大使と
                    (注) *1：「1976年7月25日」の意
                         *2：自立語辞書未収録を示す
```

図 8 自動インデクシング処理結果例

Fig. 8 Examples of output.

（c） 処理対象文は，述語で終る文を原則としている．名詞句の解析は，ロール付与上では不要であるので，「体言の修飾関係認定」で述べた以上のことは行っていない．名詞句に内包される文構造の認定は，必要性の検討を含めて，今後の課題である．

（d） 複合語の意味は，その構成単位語の単なる寄せ集めと異なる．複合語の1語としての認定は，本システムでは，辞書に1語として収録するより他はない．このため，複合語と，その末尾単位語の意味分類が異なる時，正しくロールを付与できない場合がある．

4.3 試作自動インデクシング・システムの効果

インデックス付与を人手で行うことにより，従来，実運用していた記事文検索に，本研究の自動インデクシングを適用した時の効果は，次の通りである．

（1） インデックス付与作業

（a） 従来，人手によるキーワード指定から漢字パンチを経て磁気テープに入力されるまでに要する時間は，23 分／記事であった．

（b） 一方，1記事当り平均5文（≒30 文節）の自動処理と会話型校正の結果，上記と同一の記事に対しては，ロール付きキーワードが8語程度に集約され，かつ，人手の所要時間は，5〜7 分／記事となった．これは，人手方式の 1/3 以内に改善されている．なお，平均 30 文節の記事が8語程度に集約されるのは，同一記事中で，同じ言葉がくり返し使われるからである．

（2） 検索精度

（a） 1記事当りに付与されるキーワード数は，人手方式の場合と比べて，2〜3 倍に増えると同時に，付与インデックスの標準化がなされた．この結果検索モレは，人手方式の場合より少なくなった．

（b） 人手方式の場合では，付与キーワードが少ないので，意味の広いキーワードによる検索論理式しか組めなかった．さらに，ロールによる区別ができないため，50% 程度の検索誤りが混っていた．これに対して，小規模実験の結果ではあるが，本方式では，検索誤りが 20〜30% に減少している．

（3） 外部からの情報の入手時点から，これを検索可能とする時点までの所要時間について，従来は，10〜14 日要していた．これは，キーワード抽出については，人手で逐次行うが，計算機処理については，一括処理をしていたため，人手処理と計算機処理の接続のところで待ちがあったからである．これに対して，所要時間が，従来の 10〜14 日から，4〜5 日に短縮された．

（4） その他の効果として，検索結果にロール付きキーワードを用いることにより，表形式出力や配列順指定が可能となり，編集加工性が向上した．

この結果，本研究の成果を活かした試作自動インデクシング・システムは，十分実用にたえることがわかった．

5. むすび

事実検索を指向した情報検索システムの入力情報に対し，キーワードと，情報の意味内容表示子としてのロールを自動付与することを目指して，(A)名詞と述語の依存関係に着目し，かつ，意味情報を利用した日本語文構造解析法と，(B)文構造を述語を軸に表形式にまとめた文型表を参照する方法によるプログラム構造を提案し，自動処理方式を実現した．この自動処理方式では，用語辞書照合による語の認定と，文型表参照による文構造の認定が基本技術である．また，(C)漢字プリンタ，漢字ビデオ端末を用いた会話型校正機能をも具備した具体的システムを開発し，評価した．この結果，インデックス付与作業に要する時間は，従来の人手付与方式と比べて 1/3 以内，検索精度は，検索モレを増加させることなく，検索誤り発生率を 20〜30% 低減し，また情報蓄積のターン・アラウンド時間は，1/2 以内となった．これにより，(1)情報入力蓄

積作業の省力化・標準化と，(2)検索精度の向上，の2つの目的を達成することができた．この結果，本研究の成果を活かした試作自動インデクシング・システムは，十分実用にたえることがわかった．

終りに，本研究を進めるに当り，ご討論いただいた茨城大学石綿敏雄教授，東京農工大学西村恕彦教授，ご協力いただいたファコム・ハイタック(株)清水三重二取締役，大多和英行部長，坂野匡弘課長，橋本拓主任，本研究の機会を与えていただいた(株)日立製作所システム開発研究所三浦武雄所長，本研究の遂行と本論文をまとめるに当りご指導いただいた川崎淳副所長，直接研究のご指導いただいた松岡濶主任研究員，および本システム構成プログラムを作成していただいた日立青梅電子(株)白田明氏，システム開発研究所小市健二氏に深く感謝いたします．

参考文献

1) 長尾 真ほか：日本語文献における重要語の自動抽出：情報処理 Vol. 17, No. 2 (昭 51-2).
2) 星野雅英：国文学関係論文タイトルからのキーワード自動抽出システムについて，第 14 回情報科学研究集会予稿集 (昭 52-10).
3) 石綿敏雄：日本語の生成語彙論的記述と言語処理への応用：国立国語研究所報告 54 (昭 50-4).
4) 吉田 将：2文節間の係り受けを基礎とした構文分析，信学論 Vol. 52-C, No. 10 (昭 44-10).
5) Fillmore. C.: The Case for Case, in (eds.) Bach & Harms. Universals in Linguistic Theory (New York: Hols, Rinchart & Winston) (1968).
6) 中井 浩：JICST における自然言語処理(I)，(II)，情報管理 Vol. 20, No. 11(昭 53-2), No. 12 (昭 53-3).
7) 絹川博之ほか：生活相談事例検索におけるキーワード自動抽出システム国民生活センターでのケース，情報管理 Vol. 19, No. 2 (昭 51-5).
8) 絹川博之ほか：速記反訳システム，情報処理 Vol. 16, No. 6 (昭 50-6).
9) 絹川博之ほか：自然語処理における機械辞書探索の一考察，情報処理学会第 17 回全国大会講演論文集.
10) 絹川博之ほか：日本語文構造解析による自動インデクシング方式，情報処理プログラミング・シンポジウム「日本語情報処理」報告集(昭 53-7).

(昭和54年9月5日受付)
(昭和55年2月8日採録)