

係り受け関係に基づく文献の検索

高松忍 大塚和彦
(大阪府立大学)

西田富士夫
(工学部)

1. まえがき

文献検索において、呼出し率や適合率向上のため文献の主題をいかに表現するかが重要な問題である。これまで提案されている方法⁽¹⁾⁽⁶⁾には、文献のキタームに *role-indicator* や *weight* を付加したり、キターム間を *link* したりする手法が用いられている。しかし、類似の内容をもつものが多い特許文献⁽⁸⁾ のような場合、これらの手法だけでは精密な検索ができず、より高度な主題表現法が必要である。

本考察の対象は、比較的明確な下位概念を用いて記述される特許請求範囲文などの文献要約文である。文献の主題表現として、格概念を含む係り受け構造を記述する検索言語を用い、これに文献要約文を変換する方法を与える。ここでは、従来の言語理解系⁽²⁾⁽³⁾ に比べ動詞と名詞間のより密接な依存関係に基づく解析法により係り受けのあいまいさを解消している。さらに、同じ意味内容を表わす句や節表現をあらかじめ指定した標準形の検索言語に変換する手法を与え、これにより呼出し率の向上と検索の能率化をはかる。この問題に関する従来の方法⁽⁴⁾ に比べ、適用範囲は限定されるが論理的同値関係に基づくより形式化された変換法を導入している。また、従来、単一キターム間の包含関係のみに基づいた検索が行なわれていたが⁽⁵⁾、本稿では係り受け構造を含む複合キターム間の包含関係の概念を導入し、これに基づく検索法を与える。これにより、適合率がより精密な検索を行なうことができる。

2. 検索言語と標準化

本稿が対象とする文は、(例-1) に示すような文献要約文である。

(例-1) “基板に蒸着された高抵抗率の

SiO₂層”に関する文献。

本節では、文献要約文の係り受け構造を格概念に基づいて記述する検索言語とその標準化について述べる。また、文献要約文からの検索言語への変換については3節で、標準化された検索言語を用いた文献の検索については4節で述べる。

2-1. 検索言語

一般に、文献を記述する検索言語表現は、項 t_1, t_2, \dots, t_n と格ラベル K_1, K_2, \dots, K_n により限定された関数 f を用いて、項：

$$t = f(K_1 := t_1, K_2 := t_2, \dots, K_n := t_n) \quad (1)$$

で表わす。式(1)において、項 t_i ($i=1, 2, \dots, n$) はまた式(1)のような関数表現かまたは1個の名詞の語である。 f は1個の名詞の語に相当し、その格ラベル $\{K_1, K_2, \dots, K_n\}$ は、 f が“その”、“動作を表わす名詞”、“状態を表わす名詞”等に応じてそれぞれ、 $\{ATR(\属性), MAT(材料), \dots\}$, $\{対応する動詞のとの格\}$, $\{対応する形容詞のとの格\}$ などである。これら f, t_i の種類とそれに対応する格ラベルの組を一部列挙すれば表(1)のようになる。

表(1) 標準形における f, t_i と K_i の組

P	所有述語(有る, 有, 含む)	形成の部
f のカテゴリ	その 部品 属性 材料 装置 装置・部品	
t_i のカテゴリ	属性 材料 その 部品 部品 工程 機能	
格ラベル K_i	ATTR MATR ATR2 MATZ COMP2 PROCESS FUNC	
用いる	施す	以つ 介する
	重カ	作
その	工程 状態量 場所 時間 事象 事象	
INST	MEANS COND LOC TIM PURPOSE CIRCUM	
にあり、 からで	にあり で において	に、から に、から に

P		
fのカタリ	状	態
tのカタリ	その	ものごと
格々のK _i	EXH	COMPA
	にあり	より

(注) ATRおよびMATの1,2はそれぞれ、係り受けにおける2様の現われ方を区別するためのものである。

表(1)において、Pは、関数形で表わされたことから同等な意味をもつ平叙文に対応する原子論理式の述語記号である。すなわち、前記標準形は本来ひとつの原子論理式の表わす事象を特定の項について explicit に表わした形のものと考える。自然言語表現でいえば、(例1-1)の文は“基板に蒸着された SiO₂ から成る層が高抵抗率を有する”という平叙文からえられたものと考えるのである。

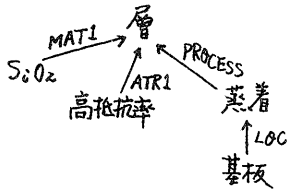
また、表(1)の後半は様相格(文献(7)参照)であり、それについては対応する各 t_i につく格助詞の語もあげてある。なお、本稿では“により”、“ために”、“において”なども格助詞相当語とみなす。以下本稿で用いる用語は文献(7)に準ずる。

(例1-2)に(例1-1)の文献要約文に対応する検索言語表現(a)を示す。また、表現(a)は(b)に示すように、関数を節点とし、格ラベルを枝にもつ木構造で記述される。

(例1-2)

(a) 層 (MAT1 := SiO₂, ATR1 := 高抵抗率, PROCESS := 蒸着 (LOC := 基板))

(b)



2-2. 標準化

一般に、同じ意味内容を異なった言語表現で表わすことができるが、このことが文献の呼出し率を低下させる原因となる。そこで、本節では同じ意味内容を表わす句あるいは節を標準形の検索言語表現に変換する手法について述べる。

本稿においては、検索のための文献要

約文は、(例1-1)のような名詞句または名詞節の形をしていとする。したがってこれらに関数表現したものは、(I)名詞句の場合(これを句表現とよぶ)

$$t = f(K_1 := t_1 - j_{k_1}, K_2 := t_2 - j_{k_2}, \dots, K_n := t_n - j_{k_n}) \quad (2)$$

(II)名詞節の場合(これを節表現とよぶ)

$$t = \{ f_{k_i}(P) (K_0 := \#_0, K_1 := t_1 - j_{k_1}, \dots, K_{i-1} := t_{i-1} - j_{k_{i-1}}, K_{i+1} := t_{i+1} - j_{k_{i+1}}, \dots, K_n := t_n - j_{k_n}) \}$$

または上式の形のものの積(Π)の形}

ここに、j_{k_i} (i=1,2,...,n)はK_iのもつ格助詞の語を示す。また、K₀はPの時制、態、使役と可能を指定する格(この格を特殊格とよぶ)からなり、#₀はこれらの格の項をまとめて示す記号であり、その具体的内容を表(2)に示す。

表(2) 特殊格

K ₀	時制 K ₀₁	態 K ₀₂	使役 K ₀₃	可能 K ₀₄
# ₀	現在過去	能動受動	非使役使役	非可能可能
	0 1	0 1	0 1	0 1

また、(II)の場合のPは式(3)の右辺の名詞節本来の形の平叙文

$$P(K_0 := \#_0, K_1 := t_1 - j_{k_1}, \dots, K_i := t_i - j_{k_i}, \dots, K_n := t_n - j_{k_n})$$

の述語記号Pである。なお、以下において格ラベルや格助詞の語を省略することがある。また、Pの引数部をまとめて#で表わす。

上記2つの形において、各 t_i (i=1, ..., n) はまた (I)あるいは(II)の形をしていすることもある。(I)(II)をまとめて中間表現とよぶ。

2-1で与えた標準形は、句表現に対応する(I)の形から格助詞の語を取り除くことによりえられる。したがって、標準形への変換は、

(i) 節表現を句表現に変換する

また、直接には (i) が行なえないような場合、

(ii) 節表現を別の節表現に変換する。によって (i) が行ないやすい形に変換することで行なわれる。

(i) 節表現と句表現間の変換

節表現:

$$f_{\text{SBJ}}(P)(t_0, \text{OBJ} := t - \text{を}) \quad (4)$$

$$f_{\text{OBJ}}(P)(t_0, \text{SBJ} := t - \text{が}) \quad (5)$$

や節表現を含む表現:

$$f(\dots, \text{MEANS} := \text{こと}(P)(t_0, \text{OBJ} := t - \text{を}) - \text{に} \text{お} \dots) \quad (6)$$

において、それぞれ表(1)を参照して、述語 P と f, t が属するカテゴリにより式(4), (5)は

$$f(K := t - \text{の}) \quad (6)$$

に式(6)は

$$f(\dots, K := t - \text{の} - \text{に} \dots) \quad (7)$$

に変換する。(例1-3)の(a)参照)

他に、節表現 $f_{k_i}(P)$ の中の f が "こと" の種の事象に属するとき、 $f_{k_i}(P)$ をサ変動詞の語 P の語幹が名詞化したサ変名詞の語で置きかえる変換がある。

((例1-3)の(b)参照)

(例1-3)

(a) 金属_{SBJ} (有する) (OBJ := 六角結晶構造 - を)

⇔ 金属 (ATR1 := 六角結晶構造 - の)

(b) こは (ド-ピン>する) (OBJ := Ge - を)

⇔ ド-ピン> (OBJ := Ge - の)

(ii) 節表現と節表現間の変換

この変換に用いる基本的規則は述語表現と関数表現との相互変換規則:

$$P(t_0, t_1, \dots, t_n) \Rightarrow t_i \in f_i(P)(t_0, t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n) \quad (8)$$

である。(文献(7))

さらに、 $i, j \geq 1$ とするとき、つぎの条件下で式(8)から式(9), (10)の関係が成立する。

条件(1) P の第 k 格と P_2 の第 k 格に同

じ項がはいっている。

条件(2) 時間的前後関係で矛盾が生じない。

$$\begin{aligned} (ii-a) & f_k(P_1)(t_0, t_1, \dots, t_{i-1}, \\ & f_j(P_2)(t_0, \dots, t_{j-1}, t_{j+1}, \dots, t_n), \\ & t_{i+1}, \dots, t_n) \\ &= f_k(P_2)(t_0, \dots, t_{j-1}, \\ & f_i(P_1)(t_0, \dots, t_{i-1}, t_{i+1}, \dots, t_n), \\ & t_{j+1}, \dots, t_n) \quad (9) \end{aligned}$$

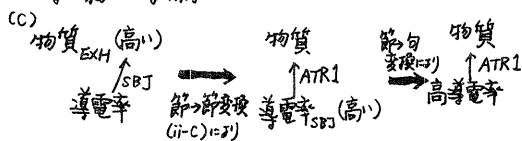
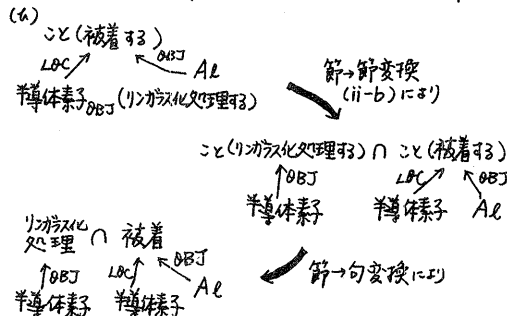
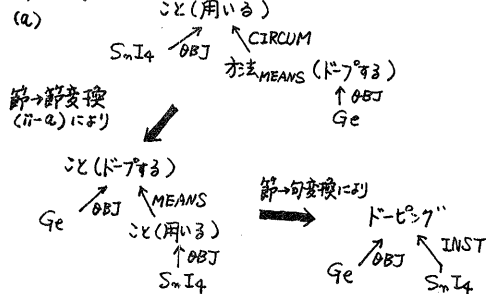
$$\begin{aligned} (ii-b) & f_k(P_1)(t_0, t_1, \dots, t_{i-1}, \\ & f_j(P_2)(t_0, \dots, t_{j-1}, t_{j+1}, \dots, t_n), \\ & t_{i+1}, \dots, t_n) \\ &= f_k(P_1)(t_0, \dots, t_{i-1}, x_i, t_{i+1}, \dots, t_n) \\ & \cap f_k(P_2)(t_0, \dots, t_{j-1}, x_i, t_{j+1}, \dots, t_n) \quad (10) \end{aligned}$$

また、述語 P が属性の程度を表わす形容詞の語で、f と t がそれぞれ、"属性" "もの" に属するとき、

$$\begin{aligned} (ii-c) & P(t_0, \text{SBJ} := f - \text{が}, \text{EXH} := t - \text{において}) \\ & \Leftrightarrow P(t_0, \text{SBJ} := f(\text{ATR2} := t - \text{の}) - \text{が}) \quad (11) \end{aligned}$$

なる関係が成立する。

(例1-4)



3. 文献要約文から検索言語への変換

入力文は、その中心事項にあたる名詞を最右端にもつ名詞句または名詞節の形の自然言語による文献要約文であるので、多様な文法的なあいまい性をもつたものがある。以下において、このような入力文を意味的な処理も含めた変換によって検索言語すなわち2節で述べた標準形に変換する手順について述べる。“文献要約文”を以下“要約文”と略称する。

3-1. 要約文の文法

要約文はまず構文解析される。このための文法を述べる前に、要約文を生成する非終端記号と前節で述べた中間表現との対応を表(3)に示す。

さて、要約文はひとつの項詞 $\langle T \rangle$ から生成され、 $\langle T \rangle$ はサ変名詞で終る句を生成する関数詞 $\langle HT \rangle$ またはそうでない名詞で終る句や節を生成する関数詞 $\langle FT \rangle$ を生成する： $\langle T \rangle \rightarrow \langle FT \rangle | \langle HT \rangle$ (12)

$\langle FT \rangle$, $\langle HT \rangle$ 以後の生成規則を 2-2. で述べた対応する中間表現および対応する要約文表現とともに表(4)に示す。

表(3) 非終端記号と中間表現の対応

非終端記号	中間表現
$\langle P(i) \rangle$ ⁽¹⁾ 述語詞(動詞, 形容詞)	P 述語記号
$\langle H \rangle$ サ変名詞	h 関数記号
$\langle F \rangle$ 変名詞以外の名詞	f
$\langle C \rangle$ 接続詞	\cap, \cup 2項論理結合子
$\langle A_0(i) \rangle$ ⁽¹⁾ 否定補助述語詞(助動詞)	\sim 否定記号
$\langle A_k(i) \rangle$ {過去受動使役, 可能 ($k=1, 2, 3, 4$)の補助述語詞(助動詞)}	"1" 特殊格 K_{0k} の項 ⁽⁵⁾
$\langle J_k \rangle, \langle J_N \rangle$ ⁽²⁾ 格助詞	j_k 格助詞の語
$\langle J_c \rangle$ 接続助詞	j_c 接続助詞の語
$\langle FT \rangle$ ⁽³⁾	{式(2)の右辺 または 式(3)の右辺の一項
$\langle HT \rangle$ ⁽⁴⁾	
$\langle T \rangle$ 項詞	t 項

注(1) $i=0, 1, 2, 3$ はそれぞれ、終止, 未然, 運用, 連体の活用を表す。
 (2) 格助詞“の”を生成する。
 (3) 中心名詞が $\langle F \rangle$ である句や節の非終端記号。
 (4) 中心名詞が $\langle H \rangle$ である句の非終端記号。
 (5) $\langle A_k(i) \rangle$ の語が存在しない場合には K_{0k} 格の項は“0”とする。

表(4) 要約文の生成

中間表現	要約文	生成規則
$f(P)(\#')$	(a) $R(3) f$	$\langle FT \rangle \rightarrow \langle P(3) \rangle \langle FT1 \rangle \langle T \rangle \langle J_k \rangle \langle FT \rangle$
$f(K_{1i}=t_{1-0}, K_{2i}=t_{2-0}, \dots, K_{ni}=t_{n-0})$	(b) $R(3) \{ t_{10}, t_{20}, \dots, t_{n0} \} f$	$\langle FT1 \rangle \rightarrow \langle T \rangle \langle J_N \rangle \langle FT1 \rangle \langle F \rangle$ $\langle F \rangle \rightarrow f, \langle T \rangle \rightarrow t, \langle J_N \rangle \rightarrow \cap$ $\langle J_k \rangle \rightarrow j_k$
$f(\dots) \cap f(P)(\#')$	(c) $R(i) j_c R'$ (R, R' は上の(a), (b) または(c)の語列を表す)	$\langle FT \rangle \rightarrow \langle P1(i) \rangle \langle J_c \rangle \langle FT \rangle$ $\langle P1(i) \rangle \rightarrow \langle T \rangle \langle J_k \rangle \langle P1(i) \rangle \langle P(i) \rangle$ $\langle P(i) \rangle \rightarrow \langle P(j) \rangle \langle A_k(i) \rangle$ $\langle J_c \rangle \rightarrow j_c$
$h(t_{1-0}, \dots, t_{i-j_{ki}})^{(注)}$ $\dots, t_{n-j_{kn}})$	(d) $\{ t_{10}, \dots, t_{i j_{ki}}, \dots, t_{n j_{kn}} \} h$	$\langle HT \rangle \rightarrow \langle T \rangle \langle J_N \rangle \langle HT \rangle$ $ \langle T \rangle \langle J_k \rangle \langle J_N \rangle \langle HT \rangle \langle H \rangle$ $\langle H \rangle \rightarrow h$
$\cap(t_1, \dots, t_{n-1}, t_n)$	(e1) $t_1, \dots, t_{n-1} \text{ と } t_n$	$\langle T \rangle \rightarrow \langle T1 \rangle$ $\langle T1 \rangle \rightarrow \langle T \rangle \langle T2 \rangle \langle C \rangle \langle T \rangle$ $\langle T2 \rangle \rightarrow \langle T2 \rangle, \langle T2 \rangle$ $\langle T2 \rangle \rightarrow \langle T \rangle$ $\langle C \rangle \rightarrow \cap$ または
$\cup(t_1, \dots, t_{n-1}, t_n)$	(e2) $t_1, \dots, t_{n-1} \text{ または } t_n$	

(注) “ j_{ki} の ” は格助詞 $\langle J_k \rangle$ の語 “ j_{ki} ” と “の” の複合助詞を示す。

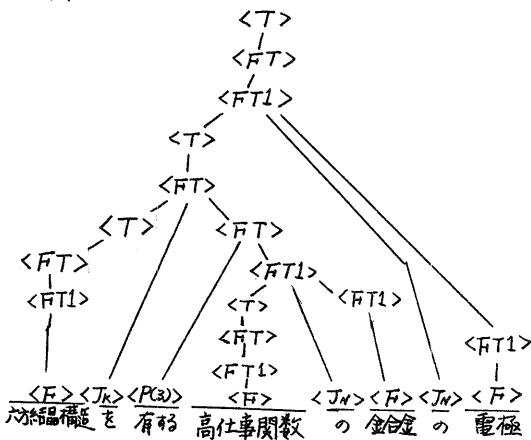
表中の $R(i)$ は平叙文に対応する中間表現が $P(i)$ である語の列:

$$\{t_1 j k_1, \dots, t_n j k_n\} P(i) [a_k(i)]$$

を示す。ここに、 $P(i)$ 、 $a_k(i)$ はそれぞれ、述語詞、補助述語詞の語を表わし、 i は表(3)におけるものと同じ意味である。また、 $\{ \}$ 中の記号列では、 $t_i j k_i$ ごとにひとまとめでしてその順序は任意であるとする。さらに、表中の小文字は、対応する大文字の表わす非終端記号から生成される端記号を示す。また、 $\langle T1 \rangle$ と $\langle T2 \rangle$ 、 $\langle FT1 \rangle$ 、 $\langle P1 \rangle$ はそれぞれ、 $\langle T \rangle$ 、 $\langle FT \rangle$ 、 $\langle P \rangle$ から生成される中間的な非終端記号である。

(例1-5) に要約文の生成例をあげる。

(例1-5)



3-2. 要約文の検索言語への変換

(a) 辞書

検索言語へ変換するために、単語辞書、分析辞書ならびに順位表を用いる。単語辞書には、単語 w を見出しとし w が属する品詞名 W を記録する。 W が名詞のときには後述する二重順位関係解消のための格間の依存関係も並記しておく。 W が述語詞のときには、述語記号の格がもつカテゴリとそのカテゴリ間の依存関係、格助詞や活用型ならびに上位述語記号を記録する。なお、サ変名詞の語の格構造は、対応するサ変動詞の語の辞書を参照する。

分析辞書には、表(4)に示すような生成規則と対応する検索言語の構成法を、順位表には、非終端記号間の順位関係を記録する。

(例1-6) に単語辞書の一部を示す。

(例1-6)

見出し語	品詞	活用型	上位(カ)	格間の依存関係
電極	<F>		部品	(成る 導電材料 SBT SOURCE) (形成形 被着 OBJ MEANS)
被着	<H>		被覆	(形成形 電極 MEANS OBJ)
有る ^(注)	<P(i)>	サ変		

(注) 述語の場合、実際のマッキングは語幹で行なう。

(a) 変換手順

検索言語すなわち標準形へ変換するための構文分析は、順位関係を主体として行ない、非終端記号間に二重順位関係が現われれば、後述する意味分析により順位を確定する。還元されるべきハンドル α とその生成規則 $A \rightarrow \alpha$ がみつかれば分析辞書の $A \rightarrow \alpha$ に対応する中間表現を構成する。構成手順は、文献(7)と同様であるので、その説明を省略し、ここでは意味分析の方法について述べる。

構文分析において、述語または関数の格への項挿入のあいまいさに起因する非終端記号間のいくつかの二重順位関係が存在する。たとえば、(例1-7)に示すようなものである。

(例1-7)

(例1-5) の文において、有る は $\langle P(3) \rangle$ に、高仕事関数 は $\langle FT1 \rangle$ にそれぞれ変換され、この $\langle P(3) \rangle$ と $\langle FT1 \rangle$ との間には

$$\dots \langle \langle P(3) \rangle \leq \langle FT1 \rangle \rangle \dots$$

なる二重順位関係が存在する。

本稿で対象とする特許請求範囲文などに対しては、項とそれが本来属すると考えられる格のカテゴリだけとの比較による従来の意味分析法によらずに上述のような係り受け関係のあいまいさが解消しない部分が多く、より専門

的な知識を用いた分析が必要である。そこで、本稿ではつぎのような原理に従って二重性を解消する。

いま、ある語の列において、その中の述語 P のある格に入りうる項が文法上では t_1, t_1' の二種類可能で、その語の列がまた t_2 なる項を含むとする。このとき述語 P と t_1, t_1', t_2 の各カテゴリ C_1, C_1', C_2 とから、

$$P(\dots, C_1, \dots, C_2, \dots)$$

または

$$P(\dots, C_1', \dots, C_2, \dots)$$

のいずれが可能であるかを辞書の格の依存関係の欄によって調べ、依存関係があれば \ni に、依存関係がなければ \ngtr に決定する。

依存関係を調べる概略の手順はつぎのようである。

単語辞書のいくつかの名詞に対しては、格の依存関係が記入されている。依存関係の欄 D が記入されている名詞 N_1' で、依存関係は、媒介となる述語 P' と相手の名詞 N_2' および N_1', N_2' が P' においてとる格の組 (P', N_2', K_1, K_2) の集合 S によって示される。したがって、名詞 N_1 と N_2 の述語 P における依存関係を調べる場合を考える。まず、 N_1 またはその上位カテゴリを表わす名詞で D が空でないものを N_1' とし、 N_1' の

依存関係の集合 S の要素で、 P', N_2' がそれぞれ P, N_2 の上位であり、格 K_1, K_2 が N_1, N_2 のそれと一致するものの集合を S_1 とする。 S_1 が1個の要素を含むならば、 N_1 は P に関して N_2 に依存関係をもつ。 S_1 が空なら N_1' のさらに上位をたどり同様のことを繰返す。なお、 S_1 が複数の要素をもつ場合は、二重順位関係は解消できない。

(例1-8)

(例1-5)の文において、高仕事関数 が $\langle FT1 \rangle$ に 有する が $\langle P(3) \rangle$ に変換されたときに生じる二重順位関係:

$$\dots \langle \langle P(3) \rangle \ni \langle FT1 \rangle \rangle \dots$$

は、六方結晶構造 \subset 物理的性質,
仕事関数 \subset 物理的性質

および

有する (SBJ:=物理的性質, OBJ:=物理的性質) が存在しないことから、

$$\dots \langle \langle P(3) \rangle \ngtr \langle FT1 \rangle \rangle \dots$$

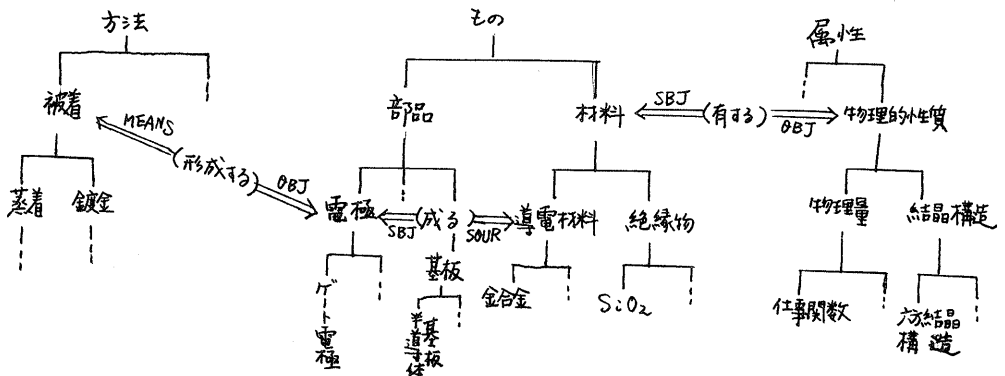
に確定し、また 高仕事関数の合金 が $\langle FT1 \rangle$ に、有する が $\langle P(3) \rangle$ に変換されたときに生じる二重順位関係:

$$\dots \langle \langle P(3) \rangle \ni \langle FT1 \rangle \rangle \dots$$

は、合金 \subset 材料, 六方結晶構造 \subset 物理的性質 および

有する (SBJ:=材料, OBJ:=物理的性質) が存在することから、

$$\dots \langle \langle P(3) \rangle \ni \langle FT1 \rangle \rangle \dots \text{に確定する。}$$



図(1) カテゴリの上位・下位関係と依存関係の例

このとき、辞書は図(1)のようになり、
 であり、 T とえば、図(1)中の
 材料 \xleftarrow{SBJ} (有する) \xrightarrow{OBJ} 物理的性質
 は
 有する(..., SBJ:=材料, ..., OBJ:=物理的性質, ...)
 なる関係を表わしているものとする。

句表現の場合の二重順位関係解消は、
 表(1)に示すような対象となっている名詞
 に関係する述語を復元することにより
 行なう。このことにより、従来より一般
 的な場合にも精密に二重順位関係の解消
 ができるが、復元可能な述語が複数個あ
 る場合がある。たとえば、

“半導体と高い仕事関数の導電材料…”

において、名詞“仕事関数”の上位カテ
 ゴリを表わす名詞“物理的性質”の依存
 関係の欄に“(を)有する”“OBJ:=物理的性質
 ”“SBJ:=材料”が見つかる。一方、“半
 導体”は“(から)成る”を介して“材料”
 に関係する。したがって、一般的には二
 重性を解消できないが、本稿における構
 文分析法では、

(1) 受け(governor)である“材料”に対
 して位置的に近い語の係り受けから分析
 する。

(2) 並列句は同じ格に入るものとする。

という方針により、

“半導体”と“高い仕事関数”

は並列句とみなさない。

上の二つの方針によって二重順位関係が
 確定できない場合には、接続詞なびう使
 用法に関する習慣などを考慮する必要が
 あるが、そのような場合の対策は現在の
 ところ行なっていない。

4. 検索

4-1. 文献の包含関係

ある文献 d の要約文の検索言語表現
 すなわち標準形を

$$t = f(K_1 := t_1, K_2 := t_2, \dots, K_m := t_m)$$

とするとき、 f, t_1, t_2, \dots, t_m の表わ
 すカテゴリを $C_0, C_1, C_2, \dots, C_m$ と
 する。

また、他の文献 d' の要約文の検索
 言語表現:

$$t' = f'(K'_1 := t'_1, K'_2 := t'_2, \dots, K'_n := t'_n)$$

における $f', t'_1, t'_2, \dots, t'_n$ の表わ
 すカテゴリを $C'_0, C'_1, C'_2, \dots, C'_n$ と
 する。

このとき、文献 d が文献 d' に包
 含されるとは、

$$C_0 \subseteq C'_0 \wedge$$

$$\forall 1 \leq n \leq n \exists 1 \leq m' \leq m' (C_m \subseteq C'_{m'} \wedge K_m = K'_{m'})$$

(13)

が成り立つこととする。また、これを
 $d \prec d'$ と表わすことにする。

(例-9)

“半導体基板に610°Cで蒸着されたSiO₂の膜”
 を d とし、

“基板への被着による絶縁物の膜”

を d' とする。このとき、検索言語表
 現はそれぞれ、

膜 (MAT1:=SiO₂, PROCESS:=蒸着 (LOC:=
 半導体基板, COND:=610°C))

膜 (MAT1:=絶縁物, PROCESS:=被着 (LOC:=基板))

であり、図(1)より、

膜 \subseteq 膜, SiO₂ \subseteq 絶縁物,

蒸着 \subseteq 被着, 半導体基板 \subseteq 基板

であるから、 $d \prec d'$ となる。

4-2 検索手順

質問 γ を記述する文の検索言語表
 現もまた文献と同様な式(1)の右辺の
 ような表現とする。検索手順は、質問
 文の検索言語表現と文献データベース中の
 検索言語表現との包含関係を調べ、質
 問が包含するデータベース中の文献を出力
 する。一般に、検索言語表現は関数を
 節点とし格ラベルを枝にもつ木構造で
 記述される。これより、式(13)に基
 づいて $d \prec \gamma$ を満たす文献 d を

求める手順は、つぎのようになる。

すなわち、従来の *tree-matching* 法と同様に、木構造の根節点から順に終端節点に向ける各枝における格ラベルの照合と各節点における関数の上位・下位関係の照合により実行する。

5. 実験結果とむすび

半導体部門の特許文献を対象として、特許請求範囲文の中間表現への変換、中間表現の検索言語への標準化ならびに文献検索手順についての小規模な計算機シミュレーションを行なった。データはリストで表わし、プログラムはアセンブラ語で書き、計算機は FACOM U-200 を用いた。以下に、入出力結果を示す。

(1) 要約文からの中間表現への変換

以下の変換過程において、記号：

$\sigma_N, \sigma_N, JK, JL, P2, P3, N1, N2, P12, P22, HN, T, C$

はそれぞれ、表(3)、(4)中の非終端記号：

$\langle F \rangle, \langle J_k \rangle, \langle J_n \rangle, \langle P(2) \rangle, \langle P(3) \rangle, \langle FT1 \rangle, \langle FT \rangle, \langle P1(2) \rangle, \langle H \rangle, \langle T \rangle, \langle C \rangle$

を示し、また、こゝで"1"式(2)、(3)の中間表現をそれぞれ、

$$f([K_1] = t_1 - j_{k_1}, [K_2] = t_2 - j_{k_2}, \dots, [K_n] = t_n - j_{k_n})$$

$$f(P)([K_1] = t_1 - j_{k_1}, \dots, [K_i] = *, \dots, [K_n] = t_n - j_{k_n})$$

と表わす。

** INPUT DATA **

入力文 # ロツホ*ウケツシヨウコウツ*ウ オ イウスル タカイ シコ*トカンスウ ノ キンコ*ウキン ノ テンキヨク @

変換
過程

(# σ_N JK P3 P3 σ_N JL σ_N JL σ_N @)

⋮

(# T JK P3 N2 JL σ_N JL σ_N @)

(# ロツホ*ウケツシヨウコウツ*ウ オ(< イウスル >([SBJ] . [σ_{BJ}])))(シコ*トカンスウ < タカイ >([SBJ] = *)) ノ キンコ*ウキン ノ テンキヨク @)

(# T JK P3 T JL σ_N JL σ_N @)

⋮

(# T JK N2 JL σ_N @)

(# ロツホ*ウケツシヨウコウツ*ウ オ(キンコ*ウキン([ATR1] = シコ*トカンスウ < タカイ >([SBJ] = *) - ノ) < イウスル >([SBJ] = * . [σ_{BJ}])) ノ テンキヨク @)

(# N2 JL σ_N @)

⋮

(# N1 @)

出力
中間表現

(#(テンキヨク([MATE1] = キンコ*ウキン([ATR1] = シコ*トカンスウ < タカイ >([SBJ] = *) - ノ) < イウスル >([SBJ] = * . [σ_{BJ}] = ロツホ*ウケツシヨウコウツ*ウ - オ) - ノ))@)

(# T @)

** INPUT DATA **

入力文 # キンオ トキンシ、ケ* イトデンキヨクオ ケイエイル コト @

出力中間表現 (#(コト < ケイエイル >([SBJ] , [OBJ] = ケ* イトデンキヨク - オ , [CIR] = * , [MEANS] = コト < トキンシル >([SBJ] , [OBJ] = キン - オ))@)

** INPUT DATA **

入力文 # ハント* ウタイノ コウセイガ*ンソ ト イオンオ コウエイル ゲ*ンソノ カコ*ウフ*ツ @

出力中間表現 (#(カコ*ウフ*ツ([MATE1] = ト(ゲ*ンソ < コウエイル >([SBJ] = * , [OBJ] = イオン - オ) , ゲ*ンソ < @ コウセイ >([SBJ] = * , [OBJ] = ハント*ウタイ - ノ)) - ノ))@)

(2) 中間表現の標準化と文献検索

以下の言語表現:

$f(K_1 t_1)(K_2 t_2) \dots (K_n t_n)$

$f(\text{REL}((K_i g P)(K_1 t_1) \dots (K_n t_n)) \dots)$

はそれぞれ、式(2), (3)の中間表現を表わす。ここに、 g は P の第 i 格 K_i のカテゴリ名である。なお、上の "REL" を含む表現をさらに標準化して検索を行なうべきであり、またそれは本稿の方法により可能であるが、検索シミュレーションでは、"REL" を含む表現で行なった。ただし、 f に係る節が1個のときには、

$((K_i f P)(K_1 t_1) \dots (K_n t_n))$

を表現した。

入力中間表現

((((SBJ フジ*ユンフ*ツト*ー七*ンク*ホウ トクチ*エトスル)(OBJ ((CIR コト モチル)(OBJ SNI4)(CIR ((MEANS ホウホウ ト*ーフ*スル)(OBJ ケ*ルマニウム)(CIR ((MEANS ホウホウ キソウセイ*イヨウスル)(SBJ (カコ*ウフ*ツハント*ウタイケツシヨウ (ATR1 ザン*ゴ*ツ*ク)))))))))))

↓ 節 → 節変換 (ii-a) により

((((SBJ フジ*ユンフ*ツト*ー七*ンク*ホウ トクチ*エトスル)(OBJ ((CIR コト ト*ーフ*スル)(MEANS ((CIR コト モチル)(OBJ SNI4)))) (OBJ ケ*ルマニウム)(CIR ((MEANS ホウホウ キソウセイ*イヨウスル)(SBJ (カコ*ウフ*ツハント*ウタイケツシヨウ (ATR1 ザン*ゴ*ツ*ク))))))))))

↓ 節 → 句変換 により

((((SBJ フジ*ユンフ*ツト*ー七*ンク*ホウ トクチ*エトスル)(OBJ ((CIR コト ト*ーフ*スル)(INST SNI4)(OBJ ケ*ルマニウム)(CIR ((MEANS ホウホウ キソウセイ*イヨウスル)(SBJ (カコ*ウフ*ツハント*ウタイケツシヨウ (ATR1 ザン*ゴ*ツ*ク))))))))))

↓ 節 → 節変換 (ii-a) により

((((SBJ フジ*ユンフ*ツト*ー七*ンク*ホウ トクチ*エトスル)(OBJ ((CIR コト キソウセイ*イヨウスル)(MEANS ((CIR コト ト*ーフ*スル)(INST SNI4)(OBJ ケ*ルマニウム))) (SBJ (カコ*ウフ*ツハント*ウタイケツシヨウ (ATR1 ザン*ゴ*ツ*ク))))))))

標準化
過程

出力検索
言語表現

質問 検索言語
表現

(QUE2)

(セイホウ(GBJ ハント*ウタイソウチ)(TOK((CIR コト ケイエイズル)(GBJ フジ*ユンフ*ツツウ)(MEANS((CIR コト カクザンズル)(GBJ フジ*ユンフ*ツ)))))

(D0C11)

(セイホウ(GBJ ハント*ウタイソウチ)(TOK((CIR コト ケイエイズル)(L0C シリコンキ*ン)(GBJ(ソウ(MATE1 ニザンカシリコン)))))(CIR コト ケイエイズル)(L0C(ソウ(MATE1 ニザンカシリコン)))))(GBJ(ソウ(MATE1 チツカシリコン)))))(CIR コト ケイエイズル)(GBJ カクザンリヨウイキ)(MEANS((CIR コト カクザンズル)(L0C シリコンキ*ン)(GBJ ガ*リウム)(MOD ヘンタクテキ)))))(CIR コト シ*ヰズル)(GBJ(ソウ(MATE1 チツカシリコン)))))(CIR コト ケイエイズル)(L0C(ソウ(MATE1 ニザンカシリコン)))))(GBJ(ソウ(MATE1 キンゾクザンカフ*ツ(ATR1 B3Z0K))))))

(D0C19)

(セイホウ(GBJ ハント*ウタイソウチ)(TOK((CIR コト ヒチヤクスル)(L0C ハント*ウタイキ*ン)(GBJ(ソウ(MATE(ハント*ウタイ(ATR タツシ コ)))))))(CIR コト ケイエイズル)(L0C(ソウ(MATE(ハント*ウタイ(ATR タツシ コ)))))(GBJ カイコウ)(MOD ヘンタクテキ))(CIR コト ケイエイズル)(GBJ(カクザンリ ヨウイキ(ATR1 キ*ヤクト*ウテンカ*タ))))(MEANS((CIR コト カクザンズル)(TRH カイコウ)(L0C(ソウ(MATE1 ハント*ウタイ)))))(GBJ(フジ*ユンフ*ツ(ATR1 キ*ヤクト*ウテンカ*タ)))))(CIR コト ケイエイズル)(L0C(ソウ(MATE ハント*ウタイ)))(GBJ テン*キヨク))))

検索された文
献の番号と
の検索言語
表現

要約文の中間表現への変換において、
以上で述べた名詞と動詞・形容詞間のよ
り密接な依存関係に基づく解析法を導入
することによりかなりの係り受けのあい
まい性を解消できた。しかしながら、本
稿で述べた方法だけでは解析不可能な部
分があり、これらはより高度なフレームの
知識ならびに与えられた文の文脈情報に
基づく解析を必要とする。また、呼出し
率向上のための中間表現の標準化におい
ても同様に、フレームの導入によるより高
度な変換の問題が残される。

なお、本検索手順では、各文献ごとに
行なうシーケンシャル方式であるが、検索の
能率化のためには文献をいくつかの標準
形ごとのグループにまとめマファイルして
おく必要がある。さらに、式(13)の条件
の意味での包含関係が成立しない場合でも
実際に包含関係があることがあり、こ
れらは今後の問題として残される。

参考文献

- (1) 安部,豊田,田中: '文献標題を利用した索引作
製の自動化に関する考察'

情報処理 Vol.11, No.12, pp.699~710 (1970)

(2) 長尾,辻米,田中: '意味および文脈情報を
用いた日本語文の解析'

情報処理 Vol.17, No.1, pp.10~28 (1976)

(3) Wilks, Y.: 'An Intelligent Analyzer
and Understander of English'
Comm. ACM Vol.18, No.5, pp.264~274 (1975)

(4) Goldman, N.M.: 'Sentence Paraphrasing
from Conceptual Base'
Comm. ACM Vol.18, No.2, pp.96~106 (1975)

(5) Turoski, W.M.: 'On a Model of
Information Retrieval System Based
on Thesaurus'
Inform. Stor. Retr. Vol.7, pp.89~94 (1971)

(6) 浅野: '日本特許情報センターの特許情報第2
検索システム'
情報管理 Vol.17, No.2, pp.127~139 (May 1974)

(7) 西田,高松: '制限された日本文の論理式への
変換と情報の抽出'
信学会論文誌 D, Vol. J59-D, No.8, pp.515~522
(1976)

(8) 特許公報 (半導体,電子回路,浜淀部門など)ら
約100件(昭48~50年のもの)