自然言語処理 45－4
（1984. 9. 28）

単語の釈義文を利用した単語間の階層関係の抽出について

鶴丸弘昭・水野浩司・内田 彰・日高 達・吉田 将
（長崎大学工学部） （九州大学工学部）

1．まえがき

近年、機械翻訳、自然言語理解など自然言語の機械処理に関する研究が盛んに行なわれるようになってきている。 自然言語の機械処理において意味処理[5]を本格的に行なうためには、実用規模の意味辞書（広い意味でのシソーラス）の開発が必要不可欠である。我々は、市販の国語辞典を高度に活用して、実用規模の意味辞書の開発に関する基礎的な研究を進めている[8],[9]。ここでの基本的な問題の一つに、単語間の階層関係をどのようにして求めるかがある。

この問題に対する一つのアプローチとして、国語辞典の語釈義文（DS：Definition Sentence ）の構造的特徴を利用して、見出し語（EW：Entry Word）に（階層）関係のある語（DW：Definition Word ）を抽出し、DWとEWとの間の階層関係を抽出するシステム（階層関係付けシステム）の開発を試みている。語釈義文の構造的な特徴として、(a) DWは語釈義文の文末に表われる場合が多い[7]、(b) DSの文末にDWとEWとの関係を規定する表現（特殊文末表現）が含まれている場合がある、などがある。

本報告は、これらの特徴を利用した階層関係付けシステムの概要とその実験結果について述べたものである。

システムの構成を図１に示す。このシステムは、次の５つの処理からなっている。

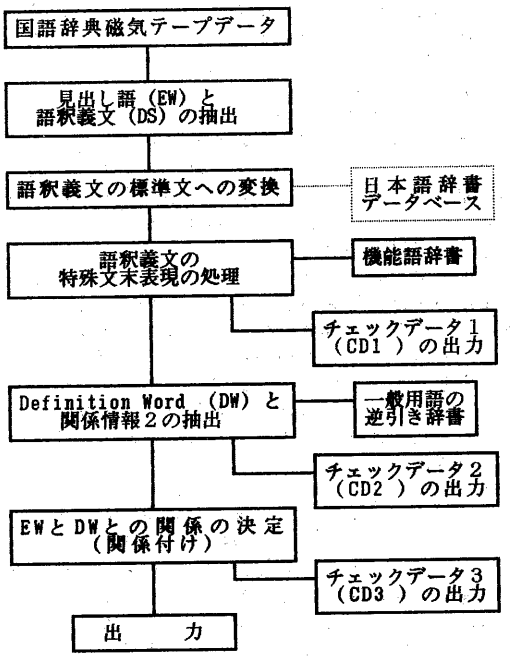

図１ システムの構成

(1) 見出し語（EW）と語釈義文（DS）の抽出。
国語辞典磁気テープデータから品詞別にEWに対応したDSを抽出する。一つのEWが複数個の語義を持っている場合があるので、それぞれの語義に対応したDSを抽出する必要がある。

(2) 語釈義文（DS）の標準文への変換。
DSには、括弧やドット（・）による省略表現があったり、漢字に読みが付加されていたりするので、DSを標準的な日本語文表現に変換しておく。
例【精鋭】：勢いがよく鋭い・こと（兵士）。
標準文 { 1.勢いがよく鋭いこと。 2.勢いがよく鋭い兵士。

(3) 語釈義文（DS）の特殊文末表現の処理。
標準文変換されたDSから、Skeleton Sentence（SS）の抽出、およびEWとDWとの関係付けに必要な関係情報１（DSの特殊文末表現から得られる情報）の抽出を行なう。SSとは、DSが特殊文末表現を含まない場合はDSそれ自身であり、DSが特殊文末表現を含む場合はDSからそれを削除した残りの部分（文）である。

(4) Definition Word （DW）と関係情報２の抽出。
(3)で求められたSSから、DWの抽出、およびEWとDWとの関係付けに必要な関係情報２（SSの構造的特徴から得られる情報）の抽出を行なう。

(5) EWとDWとの（階層）関係の決定。
(3)と(4)で求められた関係情報をもとにEWとDWとの（階層）関係の決定（関係付け）を行なう。

先に、(1)と(2)を中心に発表した[14]ので、本稿では、(3)、(4)および(5)について報告する。

なお、本研究では、三省堂の新明解国語辞典磁気テープデータ[2]を使用している。

2． 語釈義文の特殊文末表現の種類と機能

標準文変換された語釈義文約10,000文を対象に、それらに含まれている特殊文末表現を調査して、約40種の（文末）機能パターン（FP：Functional Pattern）と約 140種の（文末）機能語（FE：Functional Expression ）を求めた[13]。FPは、特殊文末表現の構造を与えるもので、FEとの組合せで、DWとEWとの関係を規定する機能を持つ。ここでは、機能パターン（FP）を表１に示す型に整理している。

機能語（FE）の例を表２に示す。FEには、DWとEWとがどのような位相（見方）での（階層）関係かを明らかにするものがある。これらは、一般用語のシソーラスにおいて有用な情報だと考えている。

特殊文末表現には、階層関係（上位－下位関係（＞）、同義関係（＝）、全体－部分関係（》））以外の関係を表わすものもある。このため、関連関係（R）を導入している。

表１．語釈義文の機能パターンの例

型	機能パターン(FP)
100	「…DW」＋（など）＋σ^*＋FE。
200	
201	…DW＋（など）＋の＋FE。
202	…DW＋（など）＋W_{202}＋の＋FE。
203	…DW＋（など）＋と＋DW＋（など）＋との＋FE。
300	
301	…DW＋（など）＋を＋σ^*＋W_{301}＋FE。
302	…DW＋（など）＋に対する＋FE。
400	…DW＋など。

注）σ^*：任意の文字列
W_{202}：意、の意、として、，、の中、の種、の一種
W_{301}：言う、いう、言った、呼ぶ、呼んだ、指す、指した

表２．機能語辞書の内容（一部）

FE	k	s	FP	rl
一つ	2	2	201	＞
			202	＞
一つ一つ	2	1	201	》
群れ	2	1	201	R
横がわ	2	1	201	＞
内がわ	2	1	201	＞
表現	2	1	100	＝＞
雅語的表現	2	1	100	＝＞
漢語的表現	2	2	100	＝＞
			201	＝＞
称	1	3	100	＝＞
			201	＝＞＜
			202	＝＞
愛称	1	1	201	＝＞
異称	1	2	100	＝＞
			201	＝＞
言い方	3	2	301	＝＞
			100	＝＞
新しい言い方	3	1	100	＝＞

注）FE：機能語，k：文末語の文字数，s：型の数
FP：機能パターンの型，rl：DWとEWとの関係情報

３．語釈義文の特殊文末表現の処理

３．１　機能語辞書

語釈義文（DS）の文末表現が機能語（FE）かどうかを判定するために、機能語辞書（FED）を用いる。この判定は、DSの文末から最長一致法で行なわれる。このため、FED は、拡張B-tree[10]の構造を持った逆引き辞書になっている。

FED は、次の内容を持っている。

(1) 逆向列の機能語（FE）
(2) FEの末尾の単語の文字数
(3) FEと組合せ可能な機能パターン（FP）の個数
(4) FEと組合せ可能なFPの型、およびそのFPが持っている関係情報１（この情報は、DWとEWとの間にどのような（階層）関係が存在するかを規定する）

表２に、FED の内容の例を示す。

３．２　Skeleton Sentence と関係情報１の抽出

標準文変換された語釈義文（DS）から、Skeleton Sentence（SS）と関係情報１の抽出を行なう。

処理手順は次の通りである。

Step 1.（機能語（FE）の判定）
　入力文（標準文変換された語釈義文（DS）、またはStep 4. からの再入力文）の文末表現がFEかどうかを、機能語辞書を用いて判定する。
　(1)入力文の文末表現がFEでなければStep 5. へ。
　(2)再入力された入力文の文末表現がFE（ただし、'など' を除く）であればStep 6. へ。

Step 2.（機能パターン（FP）の判定）
　FP判定処理ルーチンにより、入力文のFPの型を決定する。FP判定処理ルーチンは、２．の表１をもとに作成している。入力文にFPが含まれていない場合は、Step 6. へ。

Step 3.（関係情報１の抽出）
　機能語辞書よりFPに対応した関係情報１を抽出する。また、'など' が含まれていたことを関係情報１（'など' 情報）として抽出する。関係情報１が '？' であれば、Step 6. へ。

Step 4.（再入力文の生成）
　入力文からFP（または 'など'）を除去して、残りの部分を再入力文とする。Step 1. へ。

Step 5.（Skeleton Sentence （SS）の抽出）
　Step 1. の(1)の条件を満足した文をSSとして抽出する。すなわち、入力文がFPを含む場合は、入力文よりFPを削除した残りの部分がSSとなる。入力文がFEを含まない場合は、入力文がそのままSSとなる。

Step 6.（チェックデータ１（CD1）の出力）
　入力文（標準文変換されたDS）が、FEを含むがFPを含まない場合や、複数個のFEを含む場合などCD1 として、DSがそのまま出力される。このような場合、現段階ではFEとDWとの区別が機械的には困難だからである。複数個のFEの組合せとしてどのような場合があるのか、そのときDWとEWとの関

係はどのようになるのか、などを調査するための資料としての利用を考えている。

図２に、SSおよび関係情報１の抽出例を示す。図２には、SSの一部、機能語辞書（FED ）より求められた関係情報１、見出しとその正書法（EW）、語義の区別（大語義番号、小語義番号、標準文変換における通し番号）、および機能パターン（FP）の型が示されている。

図３に、CD1 の出力例を示す。図３には、図２の項目の他に、チェックの理由が示されている。

４．Definition Word （DW）と関係情報２の抽出

３．で得られたSSから、Definition Word （DW）と関係情報２の抽出を行なう。

SSからDWを抽出するために一般用語（正書法の名詞見出し約75,000語）の逆引き辞書を用いた。この逆引き辞書は参考文献(12)で作成されたものである。

DWの抽出は、基本的には、この逆引き辞書を利用した、SSの文末との最長一致法によるマッチング処理である。しかし、語釈義文（DS）には特殊な記述形式（たとえば、動植物名は片仮名表記になっているなど）があるので、それらを考慮してDWおよび関係情報２の抽出を行なった。

処理手順は次の通りである。

Step 1.（漢字列または片仮名列の抽出）
入力文（SS、またはStep 6. からの再入力文）の文末に現われる漢字列または片仮名列を抽出する。抽出された文字列の文字数をｎとする。

Step 2.（文末語の抽出）
一般用語の逆引き辞書とマッチングした、SSの文末語を抽出する。抽出された文末語の文字数をｍとする。

Step 3.（チェックデータ２（CD2 ）の出力）
ｎ＝ｍ＝０であればCD2 としてSSが出力される。すなわち、SSの文末語が、片仮名列や漢字列でなく、逆引き辞書に登録されていない場合である。

Step 4.（DWの抽出）
a. ｎ＞ｍであればStep 1. で得られた文字列をDWとする。このとき、ｍをスタックしておく。
b. ｎ≦ｍであればStep 2. で得られた文末語をDWとする。

Step 5.（関係情報２の抽出）
DWの直前に並列記号（'・' 'や' 'と'）があるかどうかを判定する。
(1) 並列記号があれば、関係情報２として 'DWが複数個ある' 情報を抽出する。と同時に、残りのDWを抽出するために、並列記号から文頭側の部分を再入力する（Step 1. へ）。

SS	rl	EW	ａｂｃ	FE	FPの型
バラ	＝＞	いばら	（０００）	雅語的表現	１００
塩化ビニール	＝＞	えんび【塩ビ】	（０００）	略	１００
盲人の官位	＞	べつとう【別当】	（０４０）	第二	２０１
羽	Ｒ	うじょう【羽状】	（００１）	形	２０１
形状・性質から見た、雲	Ｒ	うんきゅう【雲級】	（０００）	分類	２０１
嫁	＝＞	よめご【嫁御】	（０００）	尊敬語	３０２
昔の地方区画	＞	どう【道】	（０１０）	一つ	２０１
もと、生物学・鉱物学・地質学	＝＞＜	はくぶつ【博物】	（０００）	総称	２０１

注）ａ：大語義番号，ｂ：小語義番号，ｃ：標準文変換処理における通し番号，rl：関係情報１

図２．Skeleton Sentence （ss）と関係情報１の抽出例

DS	rl	EW	ａｂｃ	FE	FPの型	理由
文字を大きく二つに分けた一つ。		おんじ【音字】	（０００）	一つ		１－１
寒の期間。		かんちゅう【寒中】	（０００）	間		１－１
許された範囲の外。	？	らちがい【埒外】	（００１）	外	２０１	２－１
下の部分。	？	かぶ【下部】	（０１０）	部分	２０１	２－１
接頭語と接尾語の総称。	＝＞＜	せつじ【接辞】	（０００）	総称	２０１	２－２
生け花で、たけの低い草花類の総称。	＝＞＜	くさもの【草物】	（０００）	総称	２０１	２－２

注）ａ，ｂ，ｃ，rlは、図２と同じ

理由：１－１ …FPを満足しない
２－１ …FPを満足するが、FEの関係情報が '？'
２－２ …FPを満足するが、FEが複数個ある

図３．チェックデータ１（CD1 ）の出力例

(2) 並列記号がなければ、SSが単語（DW）かどうかを判定する。SSが単語であれば、関係情報２として'SSが単語である'情報を抽出する。

Step 6. （DWの出力）

Step 4. で抽出されたDWを出力する。

図４に、DWおよび関係情報２（関係情報１も含む）の抽出例を示す。図３には、抽出されたDW（複数個の場合は'・'で区切られている）がKWIC形式で示されている。また関係情報１、見出しとその正書法（EW）、語義の区別、機能パターン（FP）の型（これらは図２と同じ）、および関係情報２が示されている。

図５に、CD2 の出力例を示す。

５．DWとEWとの関係の決定（関係付け）

３．で求められた関係情報１、と４．で求められた関係情報２を用いて、DWとEWとの関係の決定（関係付け）を行なう。

関係付けのための条件は次の通りである。

i． DW 》 EW （全体－部分関係）
関係情報１に'》'情報があり、それだけである。

ii． DW R EW （関連関係）
関係情報１に'R'情報があり、それだけである。

iii． DW ＜ EW （下位－上位関係）
a．関係情報１に'など'情報があり、かつDSがFPを含んでいる。
b．関係情報２に'DWが複数個ある'情報があり、かつDSがFPを含んでいる。

iv． DW ＝ EW （同義関係）
a．関係情報１に'＝'情報があり、かつ関係情報２に'SSが単語である'情報がある。
b．DSとSSとDWの３つが同じ場合、EWの定義が単語でのいいかえになっている。したがって、便宜的に、同義関係に近い関係として'いいかえ関係（≧）'を導入し、DW≧EWで表わす。

v．チェックデータ３（CD3 ）
a．関係付けにあいまいさが残る。
b．SSから複数個のDWが抽出されているが、DSがFPを含んでいない。（DW＜EWにならない場合があるので）
c．DWから文頭側の５文字以内に'・'がある。（'複数個のDWがある'情報の抽出ミスをチェックするため）

vi． DW ＞ EW （上位－下位関係）
i～vの条件を満たさない。

図６に、EWとDWとの（階層）関係付けの出力例を示す。図６には、抽出されたDW（KWIC形式で示されている）、決定された（階層）関係、見出しとその正書法（EW）、語義の区別、機能語（FE）、機能パ

DW	rl	EW	a b c	FE	FPの型	d	e
雄しべ・雌しべ	＝＞＜	ずい【蕊】	（０００）	総称	２０１	4	
気体・液体	＜	えきたい【流体】	（０００）	総称	２０３	4	
股	》	うちもも【内股】	（０００）	内側	２０１		*
盲人の 官位	＞	べつとう【別当】	（０４０）	第二	２０１		
その人の犯した 犯罪	＞	ざいめい【罪名】	（０００）	名	２０１		
相手の 批評	＝＞	こうひ【高批】	（０００）	敬称	２０１		
終電車	＝＞	しゅうでん【終電】	（０００）	略	１００		*

注）a，b，c，rlは、図２と同じ　　d：'など'情報として１，'DWが複数である'情報として４を割り当てたときの和　　e：'SSが単語である'情報として＊

図４．Definition Word （DW）と関係情報２の抽出例

SS	rl	EW	a b c	FE	FPの型
相手によらず、人間愛をもってつきあう	＝＞	ゆう【友】	（０００）	漢語的表現	１００
小さなしろ・とりで	＝＞	じょうるい【城塁】	（０００）	漢語的表現	１００
しゆんば	＝＞	しゆんめ【駿馬】	（０００）	古語的表現	１００
夜分になつてから	＝＞	ばや【暮夜】	（００２）	漢語的表現	１００
夕暮れに月が出ている	＝＞	ゆうづくよ【夕月夜】	（０００）	雅語的表現	１００
いなか	＝＞	じかた【地方】	（０１０）	老人語	１００
他人・相手のむすこ	＝＞	れいそく【令息】	（０００）	敬称	２０１
ばらせん	＝＞	ばら	（０２０）	略	１００

注）a，b，c，rlは、図２と同じ

図５．チェックデータ２（CD2 ）の出力例

ターン（FP）の型、関係情報１、関係情報２が示されている。

図７に、CD3 の出力例を示す。図７には、図４の項目の他に、チェックの理由が示されている。

６．実験結果

本実験システムは、九州大学大型計算機センター FACOM M-382 、長崎大学情報処理センター FACOM M-180 II AD 上に、PL/Iおよび FORTRAN77で実現されている。 ここでは、３．の Skeleton Sentence （SS）と関係情報１の抽出、 ４．の Definition Word（DW）と関係情報２の抽出、および５．の DW とEWとの（階層）関係の決定（関係付け）、の実験結果について考察する。

実験用のデータとして、国語辞典磁気テープデータから1/20縮小辞典を作っている。1/20縮小辞典には、単語が品詞別・活用別に国語辞典の1/20の割合で含まれている。単語の品詞・活用を決めるために、京都大学の調査資料[3],[4]を利用した。この1/20縮小辞典から抽出した名詞見出し語の語釈義文のうち、標準文変換された語釈義文（DS）2,824 文を実験に用いた。

標準文変換された語釈義文（DS：2,824 文）に対して、SSが抽出されたDSは2,711 文（約96％）であり、チェックデータ（CD1 ）として出力されたDSは113 文（約４％）であった。

前段で抽出されたSS（2,711 文）に対して、文末から単語が抽出されたSSが2,502 文（約92.3％）、チェックデータ２（CD2 ）として出力された SS が 209 文（約7,7 ％）であった。前者（2,502 文）の

	DW	関係	EW	abc	FE	FPの型	rl	d	e
	顔	》	よこつつら【横っ面】	（０１０）	横がわ	２０１	》		*
	一年	》	はんとし【半年】	（００１）	半分	２０１	》		
	木	R	じゆか【樹下】	（０００）	下	２０１	R		*
ているものから出る	炎・熱・光	<	ひ【火】	（０１２）	など	４００	<	5	
	気体・液体	<	りゆうたい【流体】	（０００）	総称	２０３	<	4	
	イノシシ	=	やちよ【野猪】	（０００）	漢語的表現	２０１	=＞		*
	追試験	=	ついし【追試】	（０２０）	略	１００	=＞		*
植物の	ストツク	＞	あらせいとう	（０００）	和名	２０１	=＞		
	元素	＞	しゅうそ【臭素】	（０００）	一つ	２０１	＞		*
相手の	説	＞	こうせつ【高説】	（０００）	敬称	２０１	=＞		
	イギリス	≧	えいこく【英国】	（０００）					*
	和歌	≧	うた【歌】	（０２０）					*
	名前	≧	めいしよう【名称】	（０００）					*
運送の	料金	＞	うんちん【運賃】	（０００）					
秋の澄み渡つた	空	＞	あきぞら【秋空】	（０００）					
物に乗らず、足で歩く	こと	＞	とほ【徒歩】	（０００）					
扱い方が困難な	もの	＞	なんぶつ【難物】	（００１）					

注）a，b，c，d，e，rlは、図４と同じ

図６．階層関係システムの出力結果の例

	DW	rl	EW	abc	FE	FPの型	d	g
目的として飼う、	鶏	><	らんようしゆ【卵用種】	（００１）	品種	２０１		0
平安時代、	皇后	=><	ちゆうぐう【中宮】	（００１）	称	２０１		0
それぞれの	省	》R	しようない【省内】	（００１）	内部	２０１		0
題詠でなく作った	詩・歌		むだい【無題】	（０１０）			4	2
農業用の	機械・器具		のうきぐ【農機具】	（０００）			4	2
	正号・負号		せいふ【正負】	（０１０）			4	2
ての隣国・あたり	近所	=＞	しりん【四隣】	（００１）	漢語的表現	１００		3
に記入・計算する	ところ		こうざ【口座】	（０２０）				3
深い川・海へ移る	魚		おちうお【落（ち）魚】	（０２０）				3

注）a，b，c，d，rlは、図４と同じ

gはチェックの理由：０…関係決定にあいまいさが残る
２…FPがなく、DWが複数ある
３…DWから文頭側５文字以内に '・' がある

図７．チェックデータ３（CD3 ）の出力例

中で、DWが正しく抽出されていた SS が 2,386文（約95.4%）であった。

前段でDWが正しく抽出されたSS（2,386 文）に対して、関係情報１と関係情報２を用いて関係付けがなされたDW－EWの組が2,305 個（96.6%）、 CD3が81個（3.4 %）であった。前者（2,305 個）の中で、妥当な関係付けは2,297 個（99.6%）であった。

以下、各段階での実験結果について考察する。

（Ⅰ）Skeleton Sentence (SS)、関係情報１、チェックデータ１（CD1 ）

SSが抽出された語釈義文（DS）2,711 文に対して、⑴機能語（FE）を含まないDSが2,374 文（約87.6%）⑵FPを含んでいるDSが337 文（約12.4%）であった。

⑵の、337 文の内訳は次の通りである。

100型	214文		
200型	114文	201型	113文
		203型	１文
300型	３文	301型	２文
		302型	１文
400型	６文		

CD1 （113 文）の内訳は次の通りである。

a.FEがありFPを満足しないもの	100 文
b.FEがありFPを満足するもので、	
b-1.FEの関係情報が '？' である	10文
b-2.FEが複数個ある	３文

（Ⅱ）Definition Word (DW)、関係情報２、チェックデータ２（CD2 ）

文末語が抽出されたSS（2,502 文）に対して、⑴SSの文末語として正しい語が抽出されているSSが2,406 文（96.2%）、⑵文末語として誤った語が抽出されたSSが96文（3,8 %）であった。

⑴、⑵の内訳は次の通りである。

⑴文末語として正しい語が抽出	2,406個
ｉ．文末語がDWとして妥当	2,386個
ii．文末語がDWとして不適当	20個
a.文の記述または構造による	８個
a-1.DWとして不都合	６個
a-2.抽出したFEがDW	２個
b.標準文変換ができていない	12個
b-1.ルビが取れていない	５個
b-2.正書法が（ ）の中にある	２個
c.１個のDWしか抽出していない	５個
⑵文末語として誤った語が抽出	96個
ｉ．品詞の誤り	17個
a.見出しが名詞でない	７個
a-1.処理過程で品詞決定を誤った	２個
a-2.国語辞典中でおかしい	５個
b.見出しは名詞であるが文末語が名詞でない	４個
c.子見出しである	６個
ii．逆引き辞書の見出しの不備	72個
a.文末語が正書法で書かれているが、逆引き辞書に登録されていない	13個
a-1.平仮名書き	３個
a-2.平仮名・漢字混合	９個
a-3.片仮名・漢字混合	１個
b.文末語が正書法で書かれていない	56個
b-1.平仮名書き	44個
b-2.平仮名・漢字混合	12個
c.逆引き辞書との最長一致によるミスマッチング	３個
iii．標準文変換ができていない	３個
a.ルビが取れていない	１個
b.正書法が（ ）の中にある	１個
c.その他	１個
iv．データエラー	４個

⑴、⑵の内訳の例を次に示す。a-1.がa1. に対応している。表記は、図４の一部が省略された形式になっている。ただし、関係情報１をDWとEWとの間に入れていない。⑴のｉ．の例は、図４に示されている。

⑴のii．の例

a1. …あたる　年・月・日【甲子】(010)
a2. 備えるべき、　一定【体裁】(031) 形　201
b1.　主人と客　キャク【主客】(011)
b2.　ぐあい　具合【ぐわい】(002)
c.　高い所と低い　所【高低】(001)

⑵のｉ．の例

a1. …ことを表わ　す【逆】(200)
a2. …ように美し　い【花の】(010)
b.　…そうでな　いか【真否】(000)
c.　…ことを表わ　す【何かと言えば】(000)

⑵のii．の例

a1.　…おもち　や【おしやぶり】(000)
a2.　浮　彫り【彫上げ】(000)
a3.　塩化　ビニール【塩ビ】(000) 略 100
b1. …表わすくら　い【段位】(000)
b2.　人の顔つ　き【人相】(001)
c.　…として払　う金【場銭】(020)

⑵のiii．の例

a.　行在アンザイ　所【行宮】(000)
b.　かし・仮借【仮借】(022)
c.そば　屋一品料理〔２〕【店屋物】(002)
d.　…芸術の　行動【芸術祭】(000)

CD2 として出力されたSSは、SSの文末単語が片仮名列や漢字列でなく、逆引き辞書に登録されていないものである。

国語辞典[1]の語釈義文に用いられている単語は必ずしも漢字表記の正書法でない（平仮名書きなど）場合があるので、機械処理を困難にしている。今回、利用した逆引き辞書が正書法の単語のみから構成さ

れていることも、DWが誤って抽出されたり、SSがCD2として出力されたりする原因の一つになっていると考えられる。

(2)のｉ．のa-1.は、本システムの第１段階（見出し語とその語釈義文の抽出）での誤りである。また、(2)のｉ．のa-2.は、国語辞典[1]の記述形式のあいまいさによるもので、機械的に誤りをなくすことは困難である。

(1)のii．のb.は、本システムの第２段階（標準文変換）での誤りであり、関係情報２の抽出失敗の原因にもなっている。

（Ⅲ）DW－EW関係付け、チェックデータ３（CD3）

DWとEWとの関係付けがなされたDW－EWの組（2,317個）の内訳は次の通りである。

(1)	関係付けが妥当	2,297個
(2)	関係付けが妥当かどうか判定困難	６個
(3)	関係付けが誤り	２個

(2)、(3)のデータを次に示す。(1)の例は、図６に示されている。

(2)のデータ

歌舞伎・小説	＜	【時代物】(021)	など	400	＞＜
悲しみ・喜び	＜	【哀歓】(001)		100	＝＞
一歩の　長さ	》	【半歩】(002)	半分	201	》
離れた　都市	＞	【いなか】(012)			
き叫ぶ　習俗	＞	【哀号】(001)			
元旦の　朝	＞	【元旦】(002)			

(3)のデータ

	耳　R	【耳たぶ】(001)	部分	100	R
つづり　字	》	【字母】(011)	ーつーつ	201	》

(2)の、関係付けが妥当かどうか直感的に判定困難な場合でも、多義情報を考慮すれば妥当とみなせるものがある。(3)の、関係付けの誤りは機能語辞書の不備によるものであった。

CD3（81個）の内訳は次の通りである。

a. 関係決定にあいまいさが残る。	17個
b. 複数個のDWが抽出されているが、DSがFPを含まない。	49個
c. DWから文頭側の５文字以内に'・'がある。	15個

b.の条件は、EW＞DWの決定に利用しようとしたものもあったが、例外や判定困難な例がいくつか見出された。また、c.の条件は、'複数個のDWがある'情報の抽出ミスをチェックするためである。c.の条件で５文字で良いのかどうかなど、現在、CD3の出力データを検討している。

７．あとがき

本稿では、国語辞典の語釈義文（DS）の構造的特徴を十分に活用して、単語間の意味的関係の一つである階層関係を機械的に求める階層関係付けシステムの概要およびその実験結果について報告した。

実験結果の考察により、次のことが言える。

(1)本システムは人間支援型を前提にしているが、入力文（標準文変換されたDS）2,824文に対して、チェックデータの総数は403文（14.3%）であった。抽出プログラムの改良のため、チェックデータを検討している。

(2)DWの抽出に関して、文末語が抽出されたSS（2,502文）に対して、適切なDWが抽出されたSS（2,386文）の割合は95.4%であるが、正しい文末語が抽出されたSS（2,406文）に対しては、その割合が99.2%になる。

(3)DW－EWの関係付けにおいて、適切なDWに対して、正しく関係付けられた割合は99.6%である。DWが正しく抽出されれば、関係付けはうまくいっているとみなせる。直感的に妥当かどうか判定困難な関係付けも、多義を考慮すれば、妥当とみなせるものがある。（（Ⅲ）の(2)のデータ参照）

(4)DWとEWとの関係に、いいかえ関係（≧）を導入した。調査の範囲ではほとんどDW＝EWとみなしてよかったが、DW＞EWの可能性を考慮したためである。

(5)一般に、部分－全体関係は、a.地理的位置関係、b.体の器官や組織、c.知識学問の領域の３つの場合に限定されている[15]が、時間的関係、距離的関係も、部分－全体関係とみなしてよいと思われる例があった。（図６参照）

(6)DWとして'もの'や'こと'が抽出されている場合が多い。DWが正しく抽出されたSS（2,386文）に対して、文末が'こと'となるSSが422文（17.7%）、文末が'もの'となるSSが84文（3.5%）であった。

(7)DWの抽出およびDW－EWの関係付けの精度をあげるためには、(a)逆引き辞書の整備・補強、(b)DSの構造的特徴のより詳かい調査が必要である。

今後の研究課題として、次のようなものがある。

(1)国語辞典に含まれる全ての名詞見出し語についてDWとEWとの（階層）関係付けを行なう。

(2)階層関係にあるDWとEWとを利用して、階層構造を求める。

これについては、別の機会に報告する予定である。これは、階層構造が国語辞典の中に暗黙の形で埋め込まれているとの仮定のもとに、この階層構造を、陽な形で抽出しようとするものであり、シソーラスの自動作成の手がかりを与えるものであると考えている。この場合、単語の多義というやっかいな問題が係わってくる。

(3)語釈義文に含まれている関係語（表現）[6]の抽出と調査。

(4)語釈義文の、格情報を用いた係り受け解析[11]を行ない、意味辞書作成支援システムの開発を試みる。

本稿では、名詞について考察したが、意味辞書では、動詞、形容詞、副詞等の意味の定義[8]も必要である。

謝辞

本研究の当初より種々の助言をいただいた九州大学工学部田中武美助手、福岡大学工学部吉村賢治講師に感謝致します。

参考文献

(1) 金田一京助、金田一春彦、見坊豪紀、柴田武、山田忠雄（編）：新明解国語辞典、三省堂、第２版（1974）、第３版（1981）
(2) 横山晶一：国語辞典データベース化の準備、電子技術総合研究所彙報、Vol.41、No.11、PP.19-27（1977.11）
(3) 長尾真、辻井潤一、山上明、建部周二：国語辞典の記憶と日本語文の自動分割、情報処理、Vol.19、No.6、PP.514-521（1978.6）
(4) 長尾真：計算機による日本語文章の解析に関する研究、昭和53年度文部省科学研究費特定研究(1) 昭和53年度研究報告書（1979.2）
(5) 田中穂積：計算機による自然言語の意味処理に関する研究、電子技術総合研究所研究報告、第797号（1979.7）
(6) 首藤公昭：文節構造モデルによる日本語の機械処理に関する研究、福岡大学研究所報、第45号（1980.3）
(7) 中野洋：分類番号つけ支援システム、情報処理学会計算言語学研究会資料25-5（1981.2）
(8) S.Yoshida：CONCEPTUAL TAXONOMY FOR NATURAL LANGUAGE PROCESSING, in Kitagawa, T.(ed), Computer Science & Technologies, Ohmsha, Japan（1982）
(9) S.Yoshida,H.Tsurumaru,T.Hitaka：MAN-ASSISTED MACHINE CONSTRUCTION OF A SEMANTIC DICTIONARY FOR NATURAL LANGUEGE PROCESSING ,Proc. COLING82, PP.419-424（1982.7）
(10) 日高達、吉田将、稲永紘之：拡張B-treeによる日本語単語辞書の作成、情報処理学会自然言語処理研究会資料33-8（1982.10）
(11) 日高達、吉田将：格文法による日本語の構文解析、情報処理学会自然言語処理技術シンポジウム論文集、PP.41-PP.46（1983.6）
(12) 吉村賢治、山下明男、日高達、吉田将：専門用語の自動収集システムについて、情報処理学会自然言語処理研究会資料42-1（1984.3）
(13) 鶴丸弘昭、日高達、吉田将：国語辞典を利用したシソーラスの作成について、情報処理学会第27回（昭和58年度後期）全国大会、2H-2（1983.10）
(14) 鶴丸弘昭、内田彰、日高達、吉田将：国語辞典からの情報抽出とその構造化、情報処理学会自然言語処理研究会資料43-6（1984.5）
(15) 長尾真監修：日本語情報処理、電子通信学会（1984.5）