情報学基礎 3－1
（1986. 11. 28）

単語間の上位－下位関係の自動抽出

鶴丸弘昭　　日高達　　吉田将
（長崎大学　工学部）（九州大学　工学部）（九州工業大学）

単語間の上位－下位関係を求める問題は，大規模意味辞書（広い意味でのシソーラス）作成における最も重要な問題の一つである．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を目指した研究を進めている．

本報告は，その第一段階として，国語辞典の語義文の解析を行い，三省堂の新明解国語辞典（すでに磁気テープ）の語義文から見出し語（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（上位－下位）関係付けを行なうシステムの試作，およびその実験結果について述べたものである．

ON AUTOMATIC EXTRACTION OF HIERARCHICAL RELATION BETWEEN WORDS

Hiroaki TSURUMARU (Ngasaki University, Department of Electronics,
Bunkyo-machi 1-14, Nagasaki 852, JAPAN)
Toru HITAKA (Kyushu University 36, Department of Electronics,
Higashi-ku Hakozaki 6-10-1, Fukuoka 812, JAPAN)
Sho YOSHIDA (Faculty of Computer and Systems, Kyushu Institute of Technology,
Tobataku Sensuimachi 1-1, Kitakyushu-shi 804, JAPAN)

How to obtain hierarchical relations (e.g. superordinate-hyponym relation) between the words is one of the most important problems for constructing a practical sized semantic dictionary (thesaurus in the wide sense).

A pilot system for extracting automatically hierarchical relations from an ordinary Japanese language dictionary (in machine readable form) has been developed. By this system, the hierarchical relations among entry words in the language dictionary are to be established.

In this paper, the features of the definition sentences in the Japanese language dictionary, the outline of the pilot system, and the estimation of the experimental results are discussed.

１．まえがき

単語間の上位－下位関係を求める問題は，大規模意味辞書（広い意味でのシソーラス）作成における最も重要な問題の一つである．我々は，市販の国語辞典を高度に活用して，単語（見出し語）間の階層（上位－下位）関係付けを行うための研究を進めている(5)(9)．

国語辞典は，実用規模の単語の意味・用法に関する多くの情報を持っている．特に，その語義文は見出し語（単語）の意味を記述したものであり，一般分野の意味についての重要な情報源と考えられる．しかし，市販の国語辞典は，人の利用が前提であり，機械処理に利用することを目的に作られているわけではない．したがって，①語義文の記述は必ずしも形式化されているわけではなく，その記述には一般の日常語（自然言語）が用いられている，②記述内容をコンパクトにするために，特殊記号や省略表現が多用されている，③辞書の使用者にある程度の常識が前提とされており，使用者が当然知っていると仮定された事柄（情報）は省略されている場合がある，などが問題となる(2)．

本報告では，まず，国語辞典の語義文の特徴と構造について，次に，語義文に現れる，見出し語（Entry Word：EW）に(階層)関係のある語（Definition Word：DW）と見出し語との間の関係付けの一般的な手法について考察し，さらに，語義文からDWを抽出し，DWとEWとの間の階層（上位－下位）関係付けを行なうシステムの試作，およびその実験結果について述べている．

我々が主に用いた国語辞典は，新明解国語辞典（三省堂）(1)であり，すでに磁気テープ化されている(6)．

２．国語辞典の語義文と機能表現

２－１　語義文の特徴とその構造

語義文は見出し語(EW)の意味を言語で記述したものであるが，辞書の編者の語義の捉らえ方・記述の方針に依存した所があり，必ずしも統一した記述形式になっているとは限らない(7)．しかし，一般の市販の国語辞典では，語義記述のためのスペースに制限があるため，(a)適当な上位語（類義語・関連語）を利用して，それが表す概念をいくつかの側面から規定（制限）する，(b)同義語・言い換え語で示す，(c)いくつかの下位語で示す，などのような記述方法が取られている場合が多い．

これら上位語，類義語，同義語，言い換え語，下位語，関連語などは，単語の意味を記述（定義）する場合の中心となる語と考えられる．以下，これらの語を中心語，または定義語(Definition Word：DW)と呼び，DWと略記する．

DWが語義文のどこに記述されているのか，EWとDWとの間にどのような（階層）関係が成り立つのか，また，その関係は何を手掛に，どのようにして判定すればよいかなどの観点から，国語辞典の語義文の特徴，構造，について調査した．

語義文の特徴として，次のようなものがある．

(1) DWは，語義文の文末に現われる場合が多い．

例①：【折尺】：…たたんでしまっておける<u>ものさし</u>。

例②：【鎌】　：草を刈る<u>農機具</u>。

ここで，‘【…】’で見出し語を，‘：’以下に語義文（の文末部分）を，下線でDWを示す．以下の例でも同様である．

この例では，DWがEWの上位語（概念）になっている．

(2) 語義文の文末がDWでなく，DW（正確には，DWを修飾する語句（修飾部）も含む）とEWとの関係を規定する表現（機能表現（Functional Expression：FE）と呼ぶ）になっている場合があり，DWがこの機能表現(FE)の直前にある．

例③：【山路】：「山の中の<u>小道</u>」の意の雅語的表現。

例④：【青蛙】：…に似た，大形の<u>カエル</u>の一種。

例⑤：【嫁御】<u>嬢</u>に対する尊敬語。

例②では，DWはかぎ括弧（‘「’，‘」’）で囲まれた文（句）の末尾にきている．機能表現(FE)は，‘かぎ括弧＋の意の雅語的表現’である．この機能表現は，見出し語‘山路’と かぎ括弧で囲まれた文‘山の中の<u>小道</u>’との間の同義関係を暗黙的に規定している．この機能表現の末尾の語（‘雅語的表現’）を機能語（Functionsl Word：FW)と呼び，これが同義関係を暗黙的に表していると考えられる．なお，この例では，DWはEWの上位語になっている．

例④では，機能表現‘の一種’が，DWを含んだ文頭側の部分と見出し語との間の上位－下位関係を積極的に規定している．機能語は‘一種’である．なお，この例では，DWがEWの上位語である．

例⑤では，DWが文末表現‘に対する尊敬語’の直前にきている．この表現も機能表現の一種であり，‘尊敬語’が機能語となる．この場合，DWとEWとの間には同値関係が成り立っている．

(3) 一般に，DWは，修飾部によって，いくつかの側面から意味が制限され特殊化されている．

(4) 複数個のDWを含む場合がある．

例⑥【油脂】：<u>油</u>と<u>脂肪</u>。

例⑦【食器】：食事に使う<u>器具</u>や<u>道具</u>。

例⑧【武術】：…，<u>馬術</u>・<u>剣術</u>・<u>弓術</u>・<u>槍術</u>など。

例⑨【三軍】：<u>空軍</u>・<u>陸軍</u>・<u>海軍</u>の総称。

例⑥⑧⑨では，個々のDWがEWの下位語になっている．しかし，例⑦では，DWがEWの上位語になっている．

以上のことから，我々は，語義文を次の二つのタイプに分けている．

タイプⅠ：([修飾部] & 定義語)^{n} 。

タイプⅡ：([修飾部] & 定義語)^{n} & 機能表現。

ここで，[…] は,任意要素，‘^{n}’は，‘や’，‘と’，‘・’による（…）内の文の並列接続を表わす．すなわち，DWが複数個ある場合を示す．‘&’は，連接(Concatination)を示す．

タイプⅠは機能表現を含まない，タイプⅡは機能表現を含む，語義文の構造を表す．タイプⅡでは，機能表現を除いた残りの部分がタイプⅠと同じ構造をしている．これは語義文の基本的な構造と考えることができよう．したがって，タイプⅡの語義文から機能表現を除いた残りの部分（文または句）およびタイプⅠの語義文を基本語義文(Skeleton Sentence)と呼び，以下，SSと略記する．

機能表現は，特別に，形式化されているわけではなく，多様な表現が用いられており，しかも，辞書によって異なる場合があり得る．したがって，機械処理のためには，機能表現の性質，種類，構造を明らかにしておく必要がある．

2－2　機能表現の構造と分類

新明解国語辞典磁気テープデータからサンプルデータとして抽出した約10,000文の語義文の中の約1,500文が機能表現を含んでいるとみなせるものであった．

機能表現は多様な形態を取るが，一般に，二項関係を表す機能語を一つ含んでおり（複数個含む場合もある），これに付属語や補助用言，形式名詞などが結びついて機能表現を形成している．機能表現は，機能語の表す二項関係を基に，見出し語(EW)と基本語義文(SS)との間の階層関係（上位－下位関係（＞），同義関係（≡），全体－部分関係（《））を規定する．ただし，階層関係以外の関係を表わすこともあるので，関連関係（Ｒ）を導入している．

現在，約 170種の機能語を求めているが，機能表現の数は相当な数になるので，機能語を含む正規表現（機能パターン（Functional Pattern：FP））で機能表現を分類している．その分類を表１に示す．

表１　語義文の機能パターンの型と種類

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

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

機能表現の取る機能パターンの型は，100型と200型がほとんどで，全体の約９割以上を占めている．また，100型は新明解国語辞典に特に多用されている．300型は数も少なく特殊な記述とみなせる．他の同様な国語辞典（例えば，岩波国語辞典，新潮国語辞典など）では，ほとんどが200型である．

機能語の一部を表２に示す．機能語は，(1) EWとSSとの間の２項関係を積極的に表わしているものと，(2) EWの語用に関する情報を持っているものとに分類できる．後者（‘雅語的表現’など）は，EWとSSとの間の同義関係を暗黙に表わしていると考えることができる．語用に関する情報は，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．

これらを整理して，階層関係付けシステムで用いる機能語辞書を作成している．

表２　機能語の分類

１．関係情報を持つ			２．語用情報をもつ	
(1)上位－下位関係（＞）			・末尾が‘語’	
一種	２０１	＞	敬語	１００　≡
種類	２０１	＞		２０１　≡
(2)下位－上位関係（＜）			・末尾が‘形’	
など	４００	＜	方言形	１００　≡
	１００	＜	・末尾が‘表現’	
(3)全体－部分関係（《）			敬語表現	１００　≡
一部分	２０１	《	・末尾が‘言い方’	
横がわ	２０１	《	言い方	１００　≡
(4)関連関係		（Ｒ）		３０１　≡
・関連関係のみ			・末尾が‘称’	
変化	２０１	Ｒ	雅称	２０１　≡
	１００	Ｒ		１００　≡
・全体－部分関係			・末尾が‘名’	
上	２０１	Ｒ《	雅名	１００　≡
(5)その他			・その他	
間	２０１	？	略	１００　≡
	２０３	？		２０１　≡

３．見出し語(EW)と定義語(DW)との関係付けの手法

語義文のタイプと機能表現の特性から，次のような仮定が得られる．ここで，概念間の意味的（二項）関係をσ_{FE}で表す．σ_{FE}として，上位－下位関係，同義関係，部分－全体関係，および関連関係を考えており，それぞれを‘＞’，‘≡’，‘《’，および‘Ｒ’と表す．また，見出し語(Entry Word)をEW，語義文（Definition Sentence)をDSで表す場合がある．

(1)語義文がタイプⅠであれば，EW ≡ DSが成り立つ．

(2)語義文がタイプⅡであれば，EWσ_{FE}SSが成り立つ．

(3)([修飾部] & DW)　≦　DW

ここで，等号は左辺のDWに修飾部がない場合である．

(4) ([修飾部$_1$] & DW$_1$)$^{\#}$　≦　[修飾部$_2$] & DW$_2$

ここで，i,j＝1～n，等号は左辺が並列接続されてなく，一個のDWしか持たない場合である．

以上の仮定と，σ_{FE}の推移律により，次のようなDWとEWとの間の関係付けの条件（手順）が得られる．

（Ⅰ）語義文(DS)がタイプⅠの場合，

①DWが修飾部を持ち，

(a)DWが一個であれば，　EW＜DW

(b)DWが複数個あれば，　CD(Check Data)

②DWが修飾部を持たず，

(a)DWが一個であれば，　EW≡DW

(b)DWが複数個あれば，　EW＞DW

（Ⅱ）語義文(DS)がタイプⅡの場合，

①DWが修飾部を持ち，

(a)DWが一個であり，

σ_{FE}が‘＜’か‘≡’であれば，EW＜DW，

σ_{FE}が‘《’か‘R’であれば，EWσ_{FE}DW

そうでなければ，　CD(Check Data)

(b)DWが複数個あれば，　CD(Check Data)

②DWが修飾部を持たず，

(a)DWが一個であれば，　EWσ_{FE}DW

(b)DWが複数個であり，

σ_{FE}が‘＞’であれば，EW＞DW

そうでなければ，CD(Check Data)

CD(Check Data)は，DWとEWとの間の関係付けが矛盾する場合である．現段階では，機械的に関係を決定するための明確な基準を設けることが困難な場合であり，人の支援が必要である．

4．階層関係付けシステム[9]

4－1　階層関係付けシステムの構成

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．

(1) 見出し語（EW）と語義文（DS）の抽出．

国語辞典磁気テープデータから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（番号）に対応したDSを抽出する必要がある．

(2) 語義文（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) 語義文の機能表現の処理と関係付け情報の抽出

標準文変換されたDSから，DSの基本語義文(SS)の抽出，およびEWとDWとの関係付けに必要な関係付け情報としてDSの機能表現から得られる情報（DSのタイプおよびσ_{FE}）の抽出を行う．

(4) 定義語(DW)の抽出と関係付け情報の抽出．

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付け情報として，DWが複数個かどうか，DWに修飾部があるかどうかの情報を抽出する．

(5) EWとDWとの（階層）関係の決定．

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

(1)と(2)については既に詳しく発表しており，また新明解国語辞典の表記法の特殊性に基づくものなので，以下，(3)，(4)および(5)について説明する．

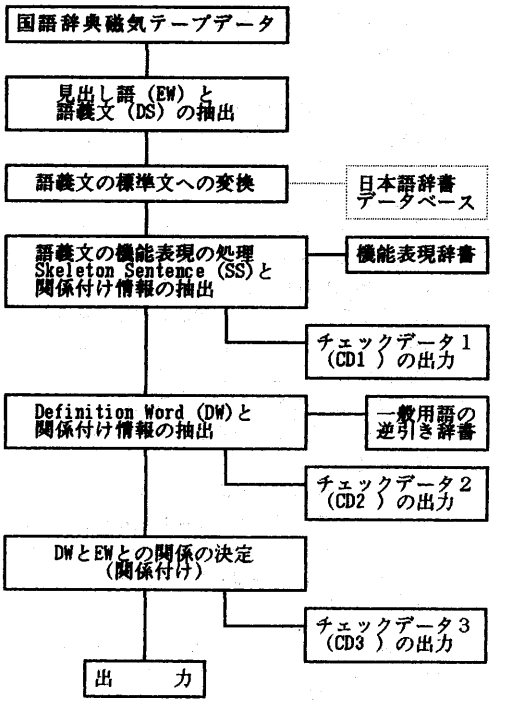

図1　システムの構成

4－2　語義文の機能表現の処理

(1)　機能語辞書

語義文（DS）の文末表現が機能語（FW）かどうかを判定するために，機能語辞書（FWD）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，FWD は，拡張B-treeの[4]構造を持った逆引き辞書になっている．

FWD は，次の内容を持っている．

(a) 逆向列の機能語（FW）

(b) FWの末尾の単語の文字数

(c) FWと組合せ可能な機能パターン（FP）の個数

(d) FW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（この情報は，SS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）

FED の内容の一部は，表2に示してある．

(2) SSの抽出と関係付け情報の抽出

標準文変換された語義文(DS)から，SSと関係付け情報の抽出を行う．処理手順は次の通りである．

Step 1. （機能語（FW）の判定）

入力文（DS，または，Step 4. からの再入力文）の文末表現がFWかどうかを，FWDを用いて判定する．

①入力文の文末表現がFWでなければ，Step 5. へ．

②再入力文の文末表現が FW（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章の表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. （関係付け情報の抽出）

機能語辞書よりFPに対応した関係情報σ_{FE}を抽出する．関係情報が ？ 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，DSがFWを含まないか，または，一個のFPを含む場合である．

Step 6. （チェックデータ１（CD1 ）の出力）

DSが，FWを含むがFPを含まない場合や，複数個のFWを含む場合など CD1として，そのDSが出力される．

このような場合，現段階ではFWとDWとの区別が機械的には困難だからである．複数個のFWが含まれる場合，それらの組合せがどうなっているのか，また，そのときDWとEWとの関係はどのようになるのか，などの調査に利用する．

図２に，SSおよび関係付け情報の抽出例を示す．図２には，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，など），および機能パターン（FP）の型などが示されている．

図３に，CD1 の出力例を示す．図３には，図２の項目の他に，チェックの理由も示されている．

４－３ DWの抽出と関係付け情報の抽出処理

(1)一般用語の逆引き辞書

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(8)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，語義文（DS）には，DWが片仮名表記（動植物名など）に変わっていたり，未知語（複合語）であったり，また，DWが複数個存在する場合などがあるので，それらを考慮してDWを抽出する必要がある．

ここで，(a) DWが他の語から修飾されてないことや，(b) 複数個のDWが抽出されたことが，EWとDWとの関係付けに必要な情報（関係付け情報）となる．

SS	α	EW	defg	FW	FPの型	β	abh
化粧をした美人	≡	ふんたい【紛黛】	（0012）	漢語的表現	100		000
屋根	R	おくじょう【屋上】	（0110）	上	201		010
塔の九輪	》	すいえん【水煙】	（0210）	一部	201		001
院とよばれるもの	R》	いんない【院内】	（0110）	内部	201		000
中年のよその男性	≡	おじさん【おじさん】	（0112）	語	301		010
使用後、処分された道具・衣服	<	ふるもの【古物】	（0011）	など	400	1	000

（注）d：大語義番号、e：小語義番号、f：文番号、g：標準文変換の際の通し番号、a：子見出し情報（0:正見出し，1:子見出し）、b：重要度（0:非重要語，1:重要語，2:最重要語）、h：標準文変換の際に取ったルビの数、α：関係情報、β：関係付け情報（1:‘など’情報，2:‘類’情報，4:DW複数情報を割り当てた時の和）

図２ ＳＳおよび関係付け情報の出力例

DS	α	EW	defg	FW	FPの型	abh	m
時局に関係して言う言葉。		じげん【時言】	（0010）	言葉		000	1－1
他人の父に対する敬称。		げんくん【厳君】	（0012）	敬称		000	1－2
上等と下等との間。	？	ちゅうとう【中等】	（0110）	間	203	000	2－1
野菜とくだもの類の総称。	≡	せいか【青果】	（0010）	総称	201	000	2－2

（注）m：チェック条件の識別情報（1-1:FWがあるがFPがない，1-2:機能表現辞書の内容の中にFPにあう型－関係情報がない，2-1:型－関係情報に‘？’がある，2-2: 201型～ 400型について、FEが複数個ある）
a～h，α：図に従う

図３ ＣＤ１の出力例

(2) DWの抽出と関係付け情報の抽出

4－2で得られたSSから，DWと関係付け情報の抽出を行なう．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，入力文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. （関係付け情報の抽出）

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，その記号から文頭側の部分を再入力する（Step 1. へ；残りのDWの抽出）．

② 並列記号がなければ，DWが他の語から修飾されていないかどうかを判定する．

Step 6. （DWの出力）

Step 4. で抽出されたDWを出力する．

図４に，DWおよび関係付け情報の抽出例を示す．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．

図５に，CD2 の出力例を示す．

４－４ DWとEWとの関係の決定（関係付け）処理

４－２で求められた関係付け情報（DSのタイプと関係情報σ_{FE}）と，４－３で求められた関係付け情報（DWが複数個あるかどうか，DWに修飾部があるかどうか）を用いて，DWとEWとの関係の決定（関係付け）を行なう．その手順は３章の条件で与えられている．３章の条件では，関係付けが一意的に決定できない場合をCDとしている．ここでは，このCDをチェックデータ３（CD3）と呼んでいる． CD3として，EWとDW（SSも合わせて）が出力される． CD3の条件を機械的に判定するために，対象とする国語辞書の語義文の特徴を利用している．たとえば，DWが修飾部を持ち，かつDWが複数個ある場合の判定として，‘DWから文頭側の５文字以内に並列接続記号‘・’がある’という条件を付加している．これは，DWが複数個あるかどうかの情報の抽出ミスをチェックするためである．

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（FW），機能パターン（FP）の型，関係付け情報が示されている．

図７に，CD3 の出力例を示す．図７には，図６の項目の他に，チェックの理由が示されている．

SSの残り	DW	α	EW	defg	FW	FPの型	β	abh	γ
男女間の	言い争い	≡	くぜつ【口舌】	（００１２）	老人語	１００		０００	
もと、	陸軍兵科〔２〕	＞	けんぺい【憲兵】	（０１１０）	一つ	２０１		０００	
	空軍・陸軍・海軍	≡	さんぐん【三軍】	（０２１０）	総称	２０１	４	０００	
使用後、処分された	道具・衣服	＜	ふるもの【古物】	（００１１）	など	４００	５	０００	
	警官		ポリス【ポリス】	（０２１０）				０００	＊
マダイの代用にする	海産硬骨魚〔２〕		あまだい【甘鯛】	（００１０）				０００	

（注）γ：関係情報（＊：SS単語情報，☆：‘・’情報）、ａ～ｈ，α～β：図に従う
DWに付いている‘〔’と‘〕’で囲まれた数（陸軍兵科〔２〕など）は、SSの文末から漢字列や片カナ列が抽出された場合に付く、一般逆引き辞書とマッチしたキーの長さ

図４ ＤＷおよび関係付け情報の出力例

SS	α	EW	defg	FW	FPの型	abh	p
将棋のこま	＞	きょうしゃ【香車】	（００１０）	一つ	２０１	０００	１
ありさま・様子	≡	たいよう【態様】	（００１０）	法律用語	１００	０００	１
足わざ		あしくせ【足癖】	（０２１０）			０００	１
生きていた時	＞R	ぞくみょう【俗名】	（０２１０）	名	２０１	０００	２

（注）ｐ：チェック条件の識別情報（1:DWが抽出されない、2:関係情報を持ち抽出されたDWが『時』だけ）
ａ～ｈ，α：図に従う

図５ ＣＤ２の出力例

SSの残り	DW	r	EW	defg	FW	FPの型	α	β	abh	γ
	音階・旋律	R	せんぽう【旋法】	(0010)	型	201	R	4	000	
	一生	》	はんせい【半生】	(0010)	半分	201	》		000	*
もと、	陸軍兵科〔2〕	>	けんぺい【憲兵】	(0010)	一つ	201	>		000	
…、大形の	竹	>	おだけ【雄竹】	(0010)	俗称	201	≡		000	
	イモリ〔0〕	≡	あかはら【赤腹】	(0310)	俗称	100	≡		000	*
	気体・液体	<	りゅうたい【流体】	(0010)	総称	203	<		000	*
	太陽・月	<	じゆげつ【日月】	(0110)					000	*
	警官	≡	ポリス【ポリス】	(0210)					000	*
日本古来の	音楽	>	ほうがく【邦楽】	(0010)					000	
赤い色の	花	>	たんか【丹花】	(0010)					000	

(注) r：DW－EW間に付けられた関係、a～h，α～γ：図に従う

図6　DW－EWとの関係付けの出力例

SSの残り	DW	EW	defg	FW	FPの型	α	β	abh	γ	s
	一家・一門	かもん【家門】	(0110)	全体	201	<》	4	000		0
食事に使う	器具・道具	しょっき【食器】	(0010)				4	010		2
他人・相手の	孫	れいそん【令孫】	(0010)	敬称	201	≡		000	☆	3
紙の	容器・紙箱	しき【紙器】	(0010)	総称	201	≡	5	000		4

(注) s：チェック条件の識別情報（0:関係付けにあいまいさが残る，1:‘類’情報がある，2:関係情報がなくDW複数情報がある，3:‘・’情報がある，4:関係付け条件‘<’を満足せず、関係情報があり、かつ、DW複数情報と‘など’情報のうち少なくとも一方がある）、a～h，α～γ：図に従う

図7　CD3の出力例

5．実験結果

本実験システムは，長崎大学情報処理センターFACOM M-180上に，主としてPL/Iで実現されている．

実験用のデータとして，標準文変換された語義文5,045文を用いた．これらは，サ変動詞や形容動詞の語幹にならない名詞見出し語の語義文約60,000文の中から任意に選らばれたものである．

表3と表4に実験の結果を示す．表3には，処理の各段階で得られた，入力データ数，出力データ数，およびチェックデータ数などが示してある．表4には，関係付けされたDWとEWとの（階層）関係の内訳が示してある．

実験の結果より，次のことがいえる．

表3　各段階での入力、出力、およびチェックデータの数

	入力	出力（妥当：誤り）	チェックデータ
(1)SSの抽出	5,045	タイプI 4,321 タイプII 527	197
(2)DWの抽出	4,848	4,516 (4,240: 276)	332
(3)関係付け	4,240	4,122 (4,115: 7)	118
実験結果	5,045	4,398 (4,115: 283)	647

表4　DW－EW関係付け出力データの内訳

関係	妥当		誤り		小計	
	タイプ		タイプ		タイプ	
DW：EW	I	II	I	II	I	II
>	3427	127	1	0	3428	127
≡	305	170	0	0	305	170
<	34	21	1	2	35	23
》	0	10	0	3	0	13
R	0	21	0	0	0	21
小計	3766	349	2	5	3768	354
計	4115		7		4122	

(1) SSが抽出された語義文（DS）4,848文に対して，
①タイプⅠの語義文(DS)が 4,321文（約89.1%）
②タイプⅡの語義文(DS)が 527文（約10.9%）
(2) 最初の入力文（標準文変換されたDS）5,045 文に対して，出力データが 4,398文（87.2%），チェックデータの総数は647 文（12.8%）であった．
①出力データ 4,398文に対して， 4,115文（93.6%）に妥当な関係付けがなされた．試作システムの解析精度は約93.6%であるといえる．
② 出力データ 4,398文に対して，283文（ 6.4%）がエラーであった．エラーのほとんどがDWの抽出で起こっている．
(4) エラーの原因として，次ぎのようなものがある．
①平仮名DWの抽出における最長一致法の限界．
②DW抽出における一般用語辞書の見出し語の不足．
③国語辞典磁気テープデータのデータエラー．
(5) 関係付けが妥当かどうか直感的に判定困難な場合，単語の多義情報（語義番号など）を考慮している．

6．あとがき

本稿では，国語辞典の語義文（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（上位－下位）関係を機械的に求めるための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．

現在，名詞見出し語約４万語について，（階層）関係付け，および，階層構造への統合のための作業を行なっている．これについては，別の機会に報告する予定である．

今後の研究課題として，次のようなものがある．
①機能表現の関係情報（σ_{FE}）の精密化
②動詞，形容詞，副詞等への本システムの応用
③語義文に含まれている関係語（表現）の抽出と調査
④意味辞書作成支援システムの開発[3]

謝辞　実験システムのプログラムの作成，資料の収集・調査等に協力を得た，長崎大学　内田　彰(現日本電気)，水野浩司(現トヨタ自動車)，井上　順君をはじめ研究室の諸氏に感謝します．

なお，本研究の一部は，文部省科学研究費特定研究「情報ドクメンテーションのための言語の研究」による．

参 考 文 献

(1) 金田一(京)，金田一(春)，見坊，柴田，山田：新明解国語辞典，三省堂，第2版(1974)，第3版(1981)
(2) 長尾真：言語辞書活用のための計算機プログラムシステムの開発と言語辞書の解析，昭和55，56年度科研費研究成果報告書（1982.2）
(3) S.Yoshida,H.Tsurumaru,T.Hitaka： MAN-ASSISTED MACHINE CONSTRUCTION OF A SEMANTIC DICTIONARY FOR NATURAL LANGUEGE PROCESSING, Proc. of COLING '82, PP.419-424 （1982.7）
(4) 日高，稲永，吉田：拡張B-treeと日本語単語辞書への応用，電学誌，Vol.16，No.4，pp.335-361（1983）
(5) 鶴丸，水野，内田，日高，吉田：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，情処学自然言語処理研資45-4（1984.9）
(6) 横山，荻野：国語辞典磁気テープのドキュメント，電総研彙，Vol.48，No.8 ，PP.672-677（1984.8）
(7) 中野洋：語義記述法の問題点，文法と意味Ⅱ（草薙，南，中野，吉田共著），第３章，PP.75-127，朝倉書店（1985.5）
(8) 吉村，日高，吉田：日本語科学技術文における専門用語の自動抽出システム，情処学論，Vol.27，No.1，(1986)
(9) H.Tsurumaru,T.Hitaka,S.Yoshida,： AN ATTEMPT TO AUTOMATIC THESAURUS CONSTRUCTION FROM AN ORDINARY JAPANESE LANGUAGE DICTIONARY, Proc. of COLING'86, PP.445-447 （1986.8）