自然言語処理 64－2
（1987. 11. 20）

語義を考慮した単語間の階層構造の抽出について

鶴丸弘昭　兵頭竜二　松崎　功　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）

本研究は，国語辞典を利用したシソーラスの自動作成に関する基礎的研究である．これまでに，見出し語とその語義文に現れる定義語との間の階層関係付けシステムを開発し，主として名詞見出し語を対象に，約７万６千組の関係付けられたデータを得ている．しかし，ここでのデータを直接用いて単語間の階層構造を求める場合に，定義語の多義や同字異音語（例，飲料：いんりょう，のみりょう）による，上位語の不要な広がりがまず問題となる．

本稿では，定義語に読みや語義番号など情報を付加するための支援システムの作成，このシステムで補強されたデータを用いて階層構造を求めるためのプログラムの作成，および，それらの実験結果と問題点について述べる．

AUTOMATIC EXTRACTION OF HIERARCHICAL STRUCTURE OF WORDS FROM THE DEFINITION SENTENCES

Hiroaki TSURUMARU, Ryuji HYOUDOU, Isao MATSUZAKI
(Nagasaki University, Department of Elactronics,
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)

This report presents an experimental approach to automatic construction of thesaurus by an ordinary Japanese language dictionary.

A system for extracting the DW(Definition Word)-EW(Entry Word) relations from the definition sentences has been developed. However these data obtained by this system lack the semantic information e.g.the meanimg(definition) number and the kana corresponding to kanji of DW, therefore a lack of these information causes unnecessary upper expansion of the hierarchical structure.

In this paper, the outline of an editor for attaching the semantic information to DW and a pilot system for organizing these data supplied by this editor into the hierarchical structure, and the estimation of the experimental results are discussed.

1．はじめに

現在，ワープロや機械翻訳システムの実用化に伴い，自然言語の機械処理が注目を集めている．このような自然言語処理において，本格的な意味解析を行なうためには，実用規模の意味辞書（広い意味でのシソーラス）が必要不可欠となる[1][2]．しかし，膨大な数の単語の意味に関する情報を，何からどのようにして求めるか，また，それらをどのような形式で構造化するか，ということが大きな問題となる[4]

我々は，国語辞典を利用したシソーラスの自動作成に関する基礎的研究として，磁気テープ化された市販の国語辞典[3]から単語の意味に関する情報を収集・整理し，その構造化に関する研究を進めている．その第一段階として，（階層）関係付けシステムの開発を行なった[1]．これは，語義文の構造的な特徴を利用して，語義文から見出し語に（階層）関係のある語（定義語）を抽出し，それらの語の間に関係付けを行なうためのシステムである．このシステムを応用して，新明解国語辞典の主として名詞見出し語を対象に，約７万６千組の関係付けられたデータを得ている．しかし，語義文中の定義語には，語義の違いを表わす語義番号や読みなどの情報が，一般に明記されていない．従って，国語辞典から直接得られるデータを用いて単語間の階層構造を求める場合に，定義語の同形異音語（例，飲料：いんりょう，のみりょう）や多義のため，上位語の不要な広がりがまず問題となる．

本稿では，定義語に読みや語義番号を付加するための支援システムの作成，このシステムを用いて補強された関係付けデータから単語間の階層構造を求めるプログラムの作成及び，その実験結果と問題点について報告する．

2．語義文の構造と関係付けの概要

2．1　語義文の構造と定義語

語義文は見出し語の意味（語義）を同じ言語で記述したものであり，単語の意味情報に関する重要な情報源である[3]．語義文は，上位語，同義語（類義語）・言い換え語，下位語など，単語の意味を記述（定義）する場合の中心となる語（以下，定義語と呼ぶ）を用いて記述されている．

例 1【折尺】：…たたんでしまっておける<u>ものさし</u>。
例 2【青蛙】：…に似た、大形の<u>カエル</u>の一種。
例 3【紺屋】：<u>染物屋</u>。
例 4【建具】：…<u>戸</u>・<u>ふすま</u>・<u>障子</u>など。

上記の例で，‘【…】’で見出し語を，‘：’以下に語義文（の末尾）を，また下線で定義語を示す．

ここで，定義語が語義文中のどこに現われるか，見出し語と定義語との階層関係を何を手掛かりに，どのようにして求めるかが重要な問題となる．

一般に定義語は，語義文の末尾に現われる場合が多い（例 1）．しかし，必ずしもそうでなく，見出し語と定義語との関係を規定する機能表現が文末に現われる場合がある（例 2）．この場合，機能表現を取り除いた残りの文または句の末尾に定義語が存在する．更に，語義文には複数の定義語が含まれる場合がある（例 4）．これらを整理して，我々は，語義文の構造を，次のような形式で近似している．

（［修飾部］＋定義語）＊ ＋［機能表現］。

ここで，［…］は任意要素，‘＋’は連接，‘＊’は‘や’，‘と’，‘・’による（…）内の並列接続を表わす．また‘（［修飾部］＋定義語）＊’のような構造をした文または句を基本的語義文と呼んでいる．

機能表現は，見出し語と基本的語義文との間の意味的関係を表わす．この意味的関係には，上位－下位関係（＞），同義関係（≡），全体－部分関係（》）などがある．これらの関係を広い意味で階層関係と呼ぶ．ただし，階層関係以外の関係とみなせるものもあるので，関連関係（Ｒ）を導入している．

また，機能表現は多様な形態を取るが，一般に，二項関係を表わす機能語を一つ含んでおり，これに付属語や補助用言，形式名詞などが結びついて形成されている．現在，この機能表現を数種のパターンに分類している．機能語としては，約 170種を求めており，見出し語と基本的語義文との間の二項関係を積極的に表わすもの（例，“一種”，“一つ”，“一部分”，“横がわ”など）と，見出し語の語用に関する情報を持っているもの（例，“雅語的表現”，“敬語”，“言い方”，“略”など）とに分類している．後者は，見出し語と基本的語義文との間の同義関係を暗黙のうちに表わしていると考えることができる．この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である．

2．2　見出し語と定義語との関係付けの条件

語義文の構造的特徴と機能表現の持つ意味的関係から，次のような仮定が得られる．これらを，関係付けの条件と呼ぶ．ここで，機能表現の持つ意味的関係をσ_Rで，語義文をDSで，基本的語義文をSSで，定義語をDWで表わす．また，基本的語義文は，末尾の定義語が修飾部で修飾された句とみなせるから，一般に複合概念を表示しているとみなしている．以下，単語（句）とそれが表わす概念とを特に区別しないで用いる．ただし，‘｛…｝’で概念を示す場合がある．

⑴DSに機能表現がない場合，　EW ≡ DS
例：｛焼酎｝ ≡ ｛…アルコール分の強い酒｝
⑵DSに機能表現がある場合，　EW σ_R SS
例：｛泡盛｝ ＜ ｛沖縄特産の、焼酎｝
⑶（［修飾部］＋DW）≦ DW
ここで，等号は左辺に修飾部がない場合である．
例：｛…の強い酒｝ ＜ ｛酒｝
⑷（［修飾部$_i$］＋DW$_i$）＊
　　≧（［修飾部$_j$］＋DW$_j$）
ここで，i,j=1 ～n ，等号はDWが一個のときである．
例：｛食べるための<u>器具</u>や<u>道具</u>｝
　　＞ ｛食べるための<u>道具</u>｝

以上の仮定と，σ_Rの推移律により，定義語と見出し語との間の関係が求められる．

ここで，定義語が語義文中に陽に示されていない場合と，定義語と見出し語との間の関係付けが一意的に決定出来ない場合は，チェックデータとして出力される．これは，機械的に関係を決定するための明確な規準を設けることが困難であり，人間支援で関係付けを行なう必要がある．

以上のような関係付けの条件を基に，我々は（階層）

関係付けシステムを作成している．このシステムで，名詞見出し約55,000語の語義文約75,000文の処理を行なった．図1に，関係付けシステムから得られた出力例を示す．さらに，3章で述べる支援システムによる処理の結果，定義語と見出し語の組が約76,000個得られた．また，関係が決定できなかったデータが約 2,000個残った．

同様な研究が，LONGMAN英々辞典を使って行なわれている[1]．

3．定義語の多義の処理

3．1　国語辞典における単語の多義

国語辞典では，読み，綴り，語義番号などの情報により，見出し語の同形異音語，同音異義語，複数の語義の区別がなされている．以下，このような情報を，単語の複数の意味，すなわち多義を区別する情報という意味で多義情報と呼ぶことにする．これに対して，語義文中の単語には，一般に，これらの情報が与えられていないため，同形異音語や同音異義語があったり，複数の語義がある場合でも，それらの区別が一般に明確になされていないことが多い．このため，（階層）関係付けシステムで得られた定義語データを基に，単語間の階層構造を求めようとする場合に，読み，綴り，語義番号などの情報の欠如による上位語の不要な広がりなどが問題となる．

従って，定義語に同形異音語や，同音異義語，多義があるとき，どの読み，どの綴り，どの語義に対応するかを，明確にしておく必要がある．このための基礎資料として，我々が研究で使用している辞書[1]で，同形異音語，同音異義語，複数の語義をもつ見出し語や定義語がどのように取り扱われているかについて調査した．

（1）同形異音語

定義語の同形異音語とは，例えば

飲料＝いんりょう，のみりょう

縁＝えにし，えん，ふち，へり，ゆかり，よすが

のように，綴りが同じで，その読みが異なる単語であり，一種の多義と考えている．

同形異音語を持つ見出し語は，表1に示すように，約 1,400語あり，これは異なり見出し語数約55,000語に対して約 3%である．しかし，関係付けシステムで得られたデータ約76,000個に対しては，約23%（約17,400個）を占める．

表1　読みに関する調査結果

読みの種類	見出し語	読みの種類	見出し語
1	50,898	4	12
2	1,271	5	5
3	93	6	2

読みを複数持つ見出し語：1,383語（2.7%）
これを定義語に持つデータ：17,442レコード（23%）

（2）仮名書き定義語と同音異義語

我々が用いた辞書では，一般に語義文中での当用漢字以外の語は平仮名書き，動植物名はカタカナ書きされている．従って，例外的に‘かしゃ（仮借）’のように漢字表記が示されているものもあるが，一般に，そうとはかぎらない．例えば，

きょうしゃ【香車】：将棋のこまの一つ。

のように，仮名書きされた語‘こま’が定義語となっ

SSの残り	DW	r	EW	defg	FW	FPの型	α	β	abh	γ
	顔	》	よこっつら【横っ面】	(0110)	横がわ	２０１	》		000	*
…中心とした	江戸川流域〔2〕	》	えど【江戸】	(0010)	一帯	２０１	》		000	
	海	R	かいてい【海底】	(0010)	底	２０１	R		000	*
	岩	R	がんとう【岩頭】	(0010)	上	２０１	R		000	*
	気体・液体	<	りゅうたい【流体】	(0010)	総称	２０３	≡	4	000	
	土地・建物	<	ふどうさん【不動産】	(0010)	など	４００	<	5	100	
	太陽・月	<	じつげつ【日月】	(0110)				4	000	
	追試験	≡	ついし【追試】	(0210)	略	１００	≡		000	*
	警官	≡	ポリス【ポリス】	(0210)					000	*
	元素	>	しゅうそ【臭素】	(0010)	一つ	２０１	>		000	*
相手の	説	>	こうせつ【高説】	(0010)	敬称	２０１	≡		000	
日本古来の	音楽	>	ほうがく【邦楽】	(0010)					000	
海	底	>	かいてい【海底】	(0010)	底	２０１	R		000	

（注）DW：定義語(Definition Word)，r：DW－EW間に付けられた関係，EW：見出し語(Entry Word)，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，FW：機能語(Functional Word)，FP：機能パターン(Functional pattern)，α：関係情報（σ_m），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，それらの総和），a：見出し情報（0:親見出し，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，γ：関係付け情報（＊：SSが単語のとき，☆：文頭側語文字以内に‘・’がある，◎：修飾部がなくてDWが複数ある），さらに‘〔’と‘〕’で囲まれた数（たとえば，‘江戸川流域〔2〕’の2）は，DWの抽出で，一般逆引き辞書でマッチしたキーの長さを示す．

図1　関係付けデータの出力例

た場合，これには，次の複数の見出し語があり，どの見出し語（の語義）であるかを区別する必要がある．

こま【こま】：…円錐形のおもちゃ。
こま【駒】：…山形にとがらせた台形の木片。
こま【齣】：映画や小説などの一区切り。

表２　語義番号による多義の数

多義の数	異り見出語	多義の数	異り見出語
1	37,314	7	41
2	12,670	8	25
3	2,423	9	9
4	978	10	6
5	259	11	2
6	98	———	———

多義を持つ見出し語：16,511語（31%）

（３）複数の語義を持つ多義語

見出しが複数個の語義を持つ場合，一般に語義番号によってその語義（意味）が区別されている．

かね【金】：(1110)金属。
：(1210)金属を取り出す鉱石。
：(1310)貨幣。

ここで，‘(defg)’は語義番号の情報であり，ｄは大語義番号，ｅは小語義番号，ｆは文番号，ｇは標準文変換における通し番号[1]である．このような見出し語は，表２に示すように約16,500語あり，見出し語約53,000語のうちの31%を占める．

一般に，見出し（読み）が同じで綴りが異なる語は，大語義番号で区別されているか，複数の見出しが立てられている．しかし，特殊な例として，次のように綴りが同じでも大語義番号で区別されている例があった．

お　(1110)【男】：「おとこ」の意の…。
(2110)【男】：…、勢いの強い方。
⋮

ところで，見出しとして，同じ読み同じ綴りをもつ単語が複数の見出しとして立てられている場合があった．これらのなかで

①品詞情報により区別が可能な見出しの場合．
例：‘付き’，‘所’，…

これらは，異なる品詞を（付き：名詞，接助）もつので，見出しの語義の区別が可能である．

②同じ品詞のため区別ができない見出しの場合．
例：‘仕舞’，‘かげろう’，…

これらは品詞が同じであるため，定義語の語義としてどの見出しの語義であるかが区別ができなくなり，問題となる．このような見出しがどのくらいあるか，何を手掛かりにすれば区別できるか，このような見出しのなかでどのくらいの語が定義語として現われるかなどについて，現在，調査を進めている．

３．２　定義語への多義情報の付加

（１）多義情報付加のための支援システムの概要

定義語の表わす語義を区別するために，多義情報（読み，綴り，語義番号など）を付加するための支援システムを作成した[1]．このシステムは図２に示すように，３つのファイルを持っている．

①処理対象ファイル
（階層）関係付けシステムで得られたデータである．

②作業用ファイル
これは，多義情報の付加，データの修正に使うために，これらの処理形態に合わせて①のファイルを変換したファイルである．

③システム用辞書ファイル
多義情報を表示するための辞書ファイルである．必要な多義情報が効率良く得られるようにするため，①のファイルを変換して使用している．

なお，この支援システムは，データエラーの修正，およびチェックデータからの定義語の抽出と関係付けの処理もできるように工夫されている．

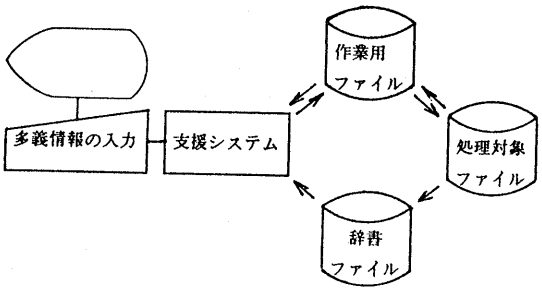


図２　支援システムの構成

（２）多義情報の付加処理の結果と問題点

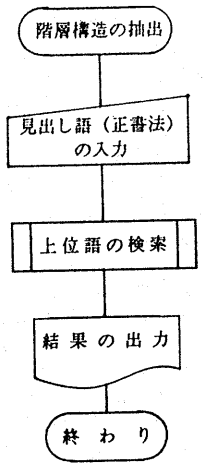
この支援システムを利用して，関係付けシステムで得られたデータ約76,000個の定義語に多義情報を付加した．この場合，語義の区別が明確でないと思われるデータがあったので，アスタリスク（＊）を導入することで，同時に複数の語義が選択できるようにした．

論理レコード長＝１４０

30B	1B	2B	1/2B	2B	1/2B	3/2B	1B	1B	1B	2*L1 B	2*L2 B	2*L3 B
EW	r	DEGF1	CODE	DEFG2	a	予備	L1	L2	L3	DW	EWの読み	DWの読み

r：定義語(DW)と見出し語(EW)の関係，a：定義語の抽出で，一般逆引き辞書でマッチした長さ
DEFG1:EWの持つ語義番号，DEFG2:DWの持つ語義番号

図３　定義語辞書のデータ形式



泡盛（あわもり）
＜：（００１０）焼酎［しょうちゅう］（００１０）
＜：（００１０）酒［さけ］（００１０）
＜：（００１０）飲料［いんりょう］（００１０）
＜：（００１０）飲物［のみもの］（００１０）
＜：（００１０）液体［えきたい］（００１１）
＜：（００１１）物質［ぶっしつ］（０１１０）
＜：（０１１０）もの［もの］　――検索不能

出力行数＝　8［行］，検索回数＝　8［回］，実行時間＝　45［ms］

器具（きぐ）
＜：（０１１０）道具［どうぐ］（０１１＊）
＜：（０１１１）器具［きぐ］　――ループの存在
＜：（０１１２）器具［きぐ］　――ループの存在
＜：（０１１３）器具［きぐ］　――ループの存在
＜：（０１１４）器具［きぐ］　――ループの存在
＜：（０２１０）器械［きかい］（００１０）
＜：（００１０）道具［どうぐ］（０１１＊）　――既に検索さえれいる

出力行数＝　8［行］，検索回数＝　3［回］，実行時間＝　27［ms］

図５　階層構造の抽出例

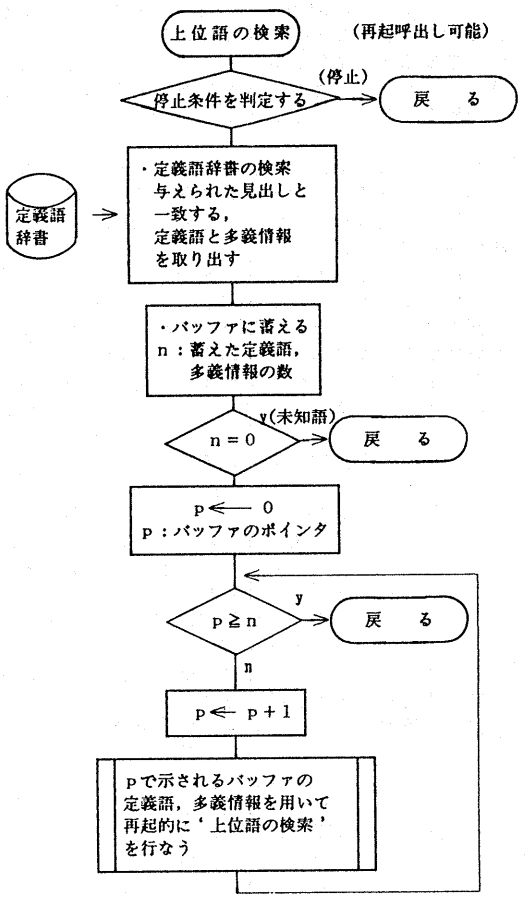


図４　階層構造抽出システムの流れ図

例えば，

ちい【地位】：社会の中で果たす役割から見た位置。

の定義語‘位置’に対応する語義は，

いち【位置】：(0111)…全体の中で占める場所。
(0112)…全体の中で占める意味。
(0211)…組織で占める、立場。
(0212)…組織で占める、境遇。
(0213)…社会で占める、立場。
(0214)…社会で占める、境遇。

があり，そのうち，(0211),(0212),(0213),(0214) のどれも正しいと思われる．このような定義語に対しては，(021*)のようにアスタリスクを入れた語義番号の情報を付加している．

ただし，アスタリスクの使用は，複数の上位語を持つことを許すことになる．このため，アスタリスクを大語義番号や小語義番号のレベルに使用した場合，階層構造を求める際に必ずしも適切な上位語だけが選択されるとは限らない場合が起こりうる．このような場合，データの検証が必要となろう．

４．多義情報を考慮した単語間の階層構造の抽出

４．１　階層構造抽出システムの概要

階層構造抽出の実験のために，３章で得られたデータを，図３に示す形式のデータに再編成している．このデータは，見出し語と，多義情報を持った定義語，およびそれらの間の関係情報から成る．チェックデータについては，定義語が入るところにその語義文の文末部分が蓄積されている．図３に示すデータから，定義語辞書を作成している．この定義語辞書は，見出し語をキーとして，その定義語と多義情報が検索できる構造になっている．

この定義語辞書を使って，単語間の階層構造を求めるプログラムを作成している[1]．このプログラムは，任意の与えられた単語をキーにして，順次定義語辞書

を検索し定義語（上位語・同義語）を求めることを繰り返し，単語間の階層構造を抽出しようとするものである．このプログラムの流れ図を図４に示す．

今回は，見出し語の上位語・同義語を順次求めることを中心に実験を行なったので，階層構造の抽出が停止する条件を，次のようにしている．

1.チェックデータである．
　チェックデータとは，語義文中に適切な定義語がないデータ，または（階層）関係が決定されていないデータのことである．
2.定義語が上位語または同義語でない．
3.定義語が，見出し語として存在しない（未知語）．
4.検索された見出し語がループを構成している．
　ある単語の上位語・同義語をたどったとき，その単語が再び現われた場合をループと呼んでいる．
5.その語が，それ以前に検索されている．
　これは，上位語の広がりを許したために生じている階層構造の冗長性である．すなわち，上位語へ行き着く経路が複数あるということになる．
6.予め設定した階層の深さの制限を越えた．
　これは，出力装置の制限などを考慮して加えられた停止条件である．

このプログラムにより求められた階層構造の出力例を図５に示す．この図において，各見出し語の後にある‘［…］’，‘(defg)’は，それぞれ定義語の読みと語義番号の情報である．

４．２　実験結果の考察と問題点

今回の実験は，３章で処理された約78,000個のデータを用いて行なった．なお，このデータで語義番号を無視した場合の異なり見出し語数は約53,000語，異り定義語数は約14,000語である．この異なり定義語のなかで約 5,500語（約39％）は，見出し語にないもの（未知語）と，見出し語にはあるがさらに上位語または同義語になる定義語をもたないものとであった．これらの語は，この段階では，さらに上位語を求めることができないので，ローカルな部分で最上位語となる．従って，ここでは，便宜上‘最上位語’と呼ぶことにする．なお，語義番号を考慮した約78,000個の全データに対して，最上位語は約29,000個に含まれていた．

これらのデータを用いた，階層構造を求める実験において，多数の最上位語や，ループのため，抽出された階層構造がぶつ切れになることが問題となる．

以下，最上位語やループの特徴について考察する．

（１）最上位語が未知語である場合

(A) 最上位語が‘こと’，‘もの’であるとき

最上位語を含む約29,000個のうち，63％の約18,000個が‘こと’を含んでおり，８％の約 2,300個が‘もの’で占められている．‘こと’を含んでいるデータの約50％は，サ変動詞の語幹になり得る名詞（以下，サ変名詞と呼ぶ）であった．サ変名詞など，定義語が‘こと’であるものは，その語義文の末尾が‘～すること’のようになっており，用言的な特色が強い．従って，‘こと’の前に出現する用言（動詞）を利用して，用言との間の関係付けを考える必要がある．

(B) 最上位語が複合語のとき

例：‘常緑高木’，‘粉たばこ’，…

このような語は，最上位語のサンプルデータの約53％（サンプルデータとは，‘こと’，‘もの’を除く最上位語を含むデータ約 8,500個から任意に選ばれた 171個のデータである）を占めている．これら複合語の語構成は，例えば，‘常緑高木’＜‘高木’，‘粉たばこ’＜‘たばこ’，などのように，末尾の単語がその複合語の上位語となる場合が多く，このような関係を機械的に求めることが可能と考えられる[3]．従って，これらのデータを関係付けデータとして登録することにより，最上位語が減らせることになる．

(C) 最上位語が用言からの転生名詞のとき

例：‘鋭さ’，‘取合い’，‘企み’，…

このような語は，最上位語のサンプルデータの約 6％を占めている．これらのなかには，用言の見出し語の派生語（例，‘鋭い’と‘鋭さ’）として，または，補足的説明（例，‘取り合う’と‘取合い’）の中に名詞形として記述してあるものもある．しかし，一般に，このような語に対しては，語義文が記述されてないので，これらを関係付けデータに登録するためには，対応する定義語や関係情報をどのようにして与えるかが問題となる．

また，(A)(B)(C) の他に，‘舟’のように，見出し‘船’の補足的説明の中に‘舟とも書く’と記載してあるだけで，見出しとしては陽に現われていないために，未知語として扱われた語もある．しかし，このようなデータに対しては，補足的説明を利用して，見出しの綴りを与えてやることで解消できる．

（２）上位語・同義語を持たない最上位語の場合

例：‘上下’，‘男女’，‘気持’，‘券’，…

このような語は，最上位語のサンプルデータの約40％を占める．

これらの語の語義文では，幾つかの下位語を挙げることにより，その意味が記述してあるため，上位語を機械的に与えることは難しい．これらのなかには，お互いに意味が反する語を組み合わせて作られた語（例：‘上下’）や二つの語が集まって初めて一つの意味を構成する語（例：‘詩歌’）などが含まれている．

（３）ループを構成している場合

(A) 外来語とその訳語との間のループのとき

例：サーチライト（サーチライト）(0010)
　　＜：探照燈［たんしょうとう］(0010)
　　＜：サーチライト［サーチライト］

例：ゼロ（ゼロ）(0310)
　　＜：零［れい］(0010)
　　＜：ゼロ［ゼロ］

この場合，ループを構成する単語は，互いに同義語・言い換え語であると考えることができる．

(B) 比較的上位の語でのループのとき

例：道具（どうぐ）(0111)
　　＜：器具［きぐ］(0110)
　　＜：道具［どうぐ］

例：場所（ばしょ）(0110)
　　≡：所［ところ］(1110)
　　＜：空間［くうかん］(0111)
　　＜：場所［ばしょ］

(C) その他のループのとき

例：案（あん）(0210)

＜：原案［げんあん］(0010)
＜：案［あん］
例：文書（ぶんしょ）(0010)
＜：書類［しょるい］(0010)
≡：書付［かきつけ］(0110)
＜：文書［ぶんしょ］

(B),(C) については，ループを構成している単語が，互いに同義語または言い換え語であるとみなせるかどうかが問題となる．現在，検討中であるが，今のところ，長さ２のループ（ループの長さ：同じ語に戻るまでに検索される語の数）約70例，長さ３のループ約 5例が見つかっている．長さ２のループに対しては，‘α＜β’，‘β＜α’であれば，‘α≡β’が一般に成立するから，‘α’と‘β’は同義関係とみなせる．また，(B) に関しては，単語の持つ意味を同じ言語を用いて説明する場合には，当然生じると考えられ，同義語とみなすことが自然であろう．

以上のことから，多数のローカルな階層構造に対する処理としては，（１）-(B)のように機械的に上位語を与えることができるもの以外は，人間の介入の基に，それらの語が階層構造に繋がるようにしなければならない．

４．３　下位語の抽出

階層構造上で，下位語にはどのような語が現われるか，その広がりはどうなっているか，同じレベルの語はお互いにどような関係になっているかなども，重要な問題となる．

現在，多数の最上位語を階層構造上へ統合するための準備を進めており，下位語の抽出に関しては，必ずしも完全とはいえないが，図５の階層構造において，各レベルに現われている単語の直接の下位語の数とその代表例を表３に示す．これは，上位語による単語の一種の分類であり，シソーラスの一部を構成していることになる．

なお，同じレベルの語がお互いにどような関係になっているかは，単語の意味分類，単語の定義とも密接に関連した問題であり，このためには，それらの単語の語義文に含まれている修飾部を解析し，定義語（上位語）がどのような側面（見方）から規定されているかを明かにする必要がある．このことに関しては，次章で，その方針，問題点について述べる．

表　下位語の調査結果

上位語	下位語の数	下位語の代表例
物質	５３	液体，ガラス，無機物，…
液体	４８	飲物，水，油，…
飲物	４	飲料，甘酒，スカッシュ
飲料	６	酒，酒類，レモネード，…
酒	７１	焼酎，清酒，銘酒，…
焼酎	３	泡盛，粕取り，酎

５．単語の意味分類（単語の定義）を目指して

（１）語義文における見方とその表記（定義語の修飾部の解析）

語義文に含まれている意味情報を単語の意味分類（単語の定義）に利用するためには，定義語の修飾部を解析し，定義語がどのような側面（見方）から規定されているかを明かにすることが重要である．

例えば，‘酒’の下位語である‘焼酎’の語義文は次のように記述されている．

：酒かす・サツマイモなどを蒸留して作った、アルコール分の強い酒。

この語義文の内容から，‘焼酎’の“材料”が‘酒かす・サツマイモなど’であり，その“製造方法”が‘それらを蒸留すること’であり，その“成分”が‘アルコール分の強い’，そのような‘酒’であるとうことがわかる．ここで，“材料”，“製造方法”，“成分”などを見方と呼んでいる．語義文のなかで，このような見方はどのような表記で記述されているであろうか？例えば，製造に関する見方の一つである“材料”に対しては，次のように，さまざまな表記方法が取られている．

“～から作る（った）（‘清酒’，‘日本酒’）”，“～を原料として作った（‘葡萄酒’）”，“～から造った（‘諸白’）”，“～で造った（‘新酒’）”

単語の意味情報を語義文から抽出するためには，見方にはどのようなものがあるか，また，それらの見方が語義文のなかでどのような表記で記述されているかを明かにしておく必要がある．現在，このような見方とその表記についての調査を進めているが，見方の表記が陽になされてない場合もある．例えば，‘～に供える酒’のように“連体形＋名詞”の形式をしている場合などである．このような場合，“使用目的”などの関係が想定されることが多いが，必ずしも機械的に明確な判定基準があるわけではなく，人の高度な判断に依らざるをえない．

なお，見方は，ある程度単語を分類しその範中で設定する必要があろう．例えば，‘生産物’の場合，使用に関する見方，製造に関する見方，構成に関する見方，性質に関する見方，量に関する見方，様相に関する見方，存在場所，存在時間に関する見方などが考えらる(2)．このような見方とその表記との関係を調査したデータの整理は，単語の意味の近似的な形式化への手掛かりを与えるものと思われる．

（２）上位語からの意味情報の伝播について

単語の意味情報としてその語義文からのみでは十分でないことは明らかである．そのために，上位語の情報を利用する．例えば，‘焼酎’の語義文からは，その“使用目的”が陽には抽出できないが，４章の図５の階層構造を利用して，‘焼酎’の上位語になる‘飲物’の語義文‘飲む液体’から，これを‘飲むための液体’と“ため”を補って考えれば，‘飲物’の基本的な“使用目的”としては‘飲むこと’が抽出できる．従って，‘飲物’の下位語である‘酒’や‘焼酎’には，その“使用目的”の一つとして‘飲むこと’が与えられているとみなせる．

ところで，上位語の意味情報が下位語の性質と矛盾する場合も当然起こり得る．例えば，‘酒’の語義文

は次のように記述されている．

：米・こうじで醸造した、わが国特有の飲料。

この文からは‘酒’の“製造方法”は‘醸造すること’であり，その“材料”は‘米・こうじ’であり，また，その“存在場所”または本来の“製造場所”が‘わが国’である，そのような‘飲料’であることがわかる．しかし，この‘酒’の定義は非常に狭い意味の定義であり，‘酒’の下位語のなかで‘焼酎’のようにこの性質を満足してないものが多い．辞書中での単語の定義は，その辞書の編集方針や記述方法に依存しているわけであるから，機械辞書への応用を考える場合は，語義文だけでなく辞書中の補足的説明や用例などからの意味に関する情報を抽出する必要がある．

なお，下位語の共通の性質は上位語（の表わす概念）で代表されていると考え，上位語の意味情報が下位語の性質と矛盾する場合，下位語の性質を優先させる．

国語辞典の意味辞書作成への応用として，単語の語義文，およびその単語の上位語の語義文に含まれている意味情報を利用した，単語の定義については，文献(5)にその基本方針が述べられている．

6．おわりに

本稿では，定義語に多義情報（読みや語義番号など）を付加するための支援システムとそのシステムにより多義情報が補強された定義語データを用いた単語間の階層構造抽出システムの概要，および主として名詞見出し語を対象にした実験結果について報告した．さらに，単語間の階層関係と単語の語義文の意味情報を利用した単語の意味分類（単語の定義）についても考察している．

なお，ここで述べた実験システムは，長崎大学情報処理センターのFACOM M-360 上に，PL/Iで実現されている．

以下，まとめと今後の課題の幾つかを示す．

（1）定義語に読み・語義番号などの多義情報を与えることにより，上位語の不要な広がりを押さえることが可能になった．しかし，多義情報が一意的に確定できない場合があり，このために階層関係に冗長性が残っている．従って，今後これら冗長性のあるデータについての調査・検討が必要である．

（2）多義情報の付与により同じ表記の単語が複数の概念として取り扱われているため，下位語の広がりについての調査・検証が必要になる．

（3）多数のローカルな最上位語で，漢字複合語の場合は，比較的容易に関係付けデータが得られるが，それ以外の最上位語をいかに階層構造に統合するかが問題となる．本稿で紹介した支援システムを拡張して利用できよう．

（4）単語間のループに関しては，ほとんどの場合，同義語または類義語とみなせた．しかし，ループをいかに階層構造に統合するかが，問題として残る．また，最上位語やループが統合された場合，さらに階層構造を調査・検証する必要があろう．

（5）単語間の関係付けを機械的に判定しているため，実際には関係が曖昧な場合も考えられる．関係の精密化が必要である(1)．

（6）今回の実験では，単語間の関係として，同義関係と上位－下位関係に限定して実験を行なっているが，部分－全体関係などの調査も必要であろう．

（7）国語辞典における，単語間の階層関係を基礎に単語の定義（意味分類）を行なうためには，語義文中の定義語の修飾部の解析をはじめ，補足的説明や用例の解析も必要となる．

（8）現在，名詞見出し語だけでなく動詞や形容詞などについても，語義文の解析，および定義語の抽出や見出し語と定義語との関係付けの実験を進めている．

謝辞　実験システムによる資料の収集・整理などの作業に協力を得た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（昭49），第３版（昭56）

(2) 栗原，吉田，鶴丸，藤田：言語と思考のシミュレーション，情報社会科学講座，４，学習研究社（昭52-05）

(3)中野　洋：分類番号付け支援システム，情報処理学会計算言語学研究会資料25-5（昭56-02）

(4) 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（昭57-02）

(5) S.Yoshida,H.Tsurumaru,T.Hitaka:MAN-ASSISTED MACHINE CONSTRUCTION OF A SEMANTIC DICTIONARY FOR NATURAL LANGUAGE PROCESSING, Proc.of COLING '82, PP.419-424 (July 1982)

(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（昭59-08）

(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（昭60）

(8) 辻井潤一：辞書の構成と機械翻訳，情報処理，Vol.26，No.10，PP.1174-1183（昭60-10）

(9) 吉田　将：辞書構築における諸問題，情報処理，Vol.28，No.8，PP.933-939（昭61-08）

(10)鶴丸，日高，吉田：単語間の上位－下位関係の自動抽出，情報学基礎３－１（昭61-11）

(11)中村，酒井，長尾：英々辞典を用いた名詞の意味関係の分析，信学論NLC86-23（昭62-03）

(12)鶴丸，兵頭，松崎，日高，吉田：国語辞典からの単語間の階層関係抽出のための支援システム，第40回電気関係九支連大819（昭62-10）

(13)鶴丸，兵頭，松崎，日高，吉田：国語辞典を用いた単語間の階層構造の自動抽出，第40回電気関係九支連大820（昭62-10）

(14)荻野綱男：シソーラス作成の問題点，日本語学，Vol.6，No.5，PP.64-71（昭62-05）