自然言語処理 83－16
（1991. 5. 17）

国語辞典情報を用いたシソーラスの作成について

鶴丸弘昭　竹下克典　伊丹克企　柳川俊英　吉田　将
長崎大学工学部　九州工業大学情報工学部

シソーラスの構築では，単語間の階層関係をいかに求めておくかが最も重要な問題の一つとなる．国語辞典の語義文は，見出し語の意味を言語で記述したもので，記述をコンパクトにするために，見出し語の上位語(定義語と呼ぶ)が用いられている場合が多い．本報告では，まず，語義文の構造的特徴，および見出し語と定義語との階層関係の抽出について概説する．次に，抽出された見出し語と定義語との関係付けデータに基づくシソーラスの作成システムについて紹介し，新明解国語辞典ＭＲＤデータを利用して試作されたシソーラスにおける極大語，複数パスなど単語間の階層構造における問題点などについて考察する．

An Approach to Thesaurus Construction from Japanese Language Dictionary

Hiroaki TSURUMARU, Katsunori TAKESITA, Katsuki ITAMI, Toshihide YANAGAWA and Sho YOSHIDA*

Department of Electrical Engineering and Computer Science, Faculty of Engineering, Nagasaki University, 1-14 Bunkyou-machi, Nagasaki-shi, Nagasaki 852, JAPAN

*Department of Artificial Intelligence, Faculty of Computer Science and System Engineering, Kyushu Institute of Technology, 680-4 ooaza Kawazu, Iizuka-shi, Fukuoka 820, JAPAN

How to obtain hierarchical relations(e.g. superordinate-hyponym relation, synonym relation) is one of the most important problems for thesaurus construction. A system for extracting automatically these relations from a machine readable Japanese language dictionary(MRD) has been developed. This report introduces the principles of the mechanical extraction of the hierarchical relations from the definition sentences briefly, and deals with a trial system for constructing a thesaurus by the hierarchical relations extracted from the MRD, and discusses the problems of isolated words and multi-pathes on this thesaurus.

1．はじめに

高度な自然言語処理を目指すためには，実用規模の単語の意味に関する情報が必要となる．このような情報の収集・整理，構造化が重要な問題となっている．

国語辞典には，単語の意味・用法に関する膨大な情報が含まれている．これらは言語(語彙)に関する貴重な知識であり，特に，語義文は，見出し語の意味を言語で記述したもので，単語の意味に関する重要な情報源と考えられる．

市販の国語辞典では，語義文の記述をコンパクトにするために，見出し語の上位語を利用して，それの表す意味(概念)をいくつかの側面から限定するという方法が取られている場合が多い．このような上位語は，語義文中で意味の中心となる語と考えられるので，定義語と呼んでいる．

我々は，語義文のこのような特徴に着目し，語義文から単語の意味に関する情報，即ち，定義語や，定義語と見出し語との意味的関係など，を抽出するためのシステムを作成し，これまでに新明解国語辞典(第２版)ＭＲＤを利用してその全名詞見出し語について関係付けデータを得ている．

本報告は，これらの成果を基にしたシソーラスの試作システムについて述べたものであり，関係付けシステムで得られたデータによるシソーラスの検証，ならびに，このシソーラス上での単語間の階層構造の問題点などについての考察をしている．

2．語義文からの階層関係の抽出

2．1　語義文の構造的特徴と階層関係抽出原理

（1）語義文の構造的特徴

語義文の構造を次の二つのタイプに分類している．

タイプⅠ：([修飾部]＋定義語)*。

例①【折尺】…たたんでしまっておけるものさし。

例②【油脂】油と脂肪。

タイプⅡ：([修飾部]＋定義語)*＋[機能表現]。

例③【青蛙】…に似た、大形のカエルの一種。

例④【原爆】原子爆弾の略。

ここで，[…] は，任意要素，‘+’ は，連接を示す．‘*’ は，‘や’，‘と’，‘・’ などによる定義語の並列接続を表わす．即ち，複数の定義語がある場合を示す．また，([修飾部]＋定義語)* の構造をした文(または，句)を基本的語義文と呼ぶ．即ち，基本的語義文の末尾に定義語が現われることになる．なお，語義が複数の文で記述されている場合は，その第一文を主として対象にしている．

上記の例では，‘【…】’ の中が見出し語，その後が語義文(の文末部分)であり，定義語と機能表現が，それぞれ，下線と網掛けで示してある．

（2）見出し語と(基本的)語義文との意味的関係

[1] タイプⅠの語義文の場合，

語義文(DS)が機能表現を含まない場合であり，これ自体が基本的語義文(SS)である．見出し語(EW)の意味を記述したものであり，近似的に見出し語の表わす概念と同義な概念を表わしていると仮定する．

ＥＷ　≡　ＳＳ(　＝　ＤＳ)

例：折尺　≡　…でしまっておけるものさし

例：油脂　≡　油と脂肪

[2] タイプⅡの語義文の場合，

語義文が機能表現を含む場合である．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規定するものであり，上位－下位関係(“の一種”)や同義関係(“の略”)のほかに，全体－部分関係(“の一部分”)，集合－要素関係(“の群”)などを表すものもある．機能表現の表わす意味的関係を ρ_{FE} とすれば，次の関係が成り立つと仮定できる．

ＥＷ ρ_{FE} ＳＳ

例：青蛙　＜　…に似た、大形のカエル

例：原爆　≡　原子爆弾

（3）定義語と語義文との意味的関係

[1] 語義文に含まれる定義語(DW)が一個で，それが修飾部で修飾されている場合，次の関係が成り立つ．

修飾部＋DW　＜　DW

例：…、大形のカエル　＜　カエル

[2] 語義文に含まれる定義語が複数個で，それが修飾部で修飾されている場合，次の関係が成り立つ．

① 修飾部が一個の場合，

修飾部＋(DW)*　≡　(修飾部＋DW)*

例：食事に使う器具や道具

≡食事に使う器具や食事に使う道具

② (修飾部＋DW)が複数個の場合，

(修飾部＋DW)*　≧　修飾部＋DW

例：食事に使う器具や食事に使う道具

＞食事に使う器具

（4）見出し語と定義語との階層関係付け

以上の仮定と階層関係の推移律に基づき，見出し語と定義語との間に次のような意味的関係が成り立

つと仮定できる.

[1]語義文がタイプ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

[2]語義文がタイプⅡの場合,

①DWが修飾部を持ち,

(a)DWが一個であり,

ρ_{FE}が‘＜’か‘≡’であれば, EW ＜ DW

ρ_{FE}が‘＞’であれば, CD(Check Data)

そうでなければ, EW ρ_{FE} 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 ρ_{FE} DW

(b)DWが複数個であり,

ρ_{FE}が‘＞’か‘≡’であれば, EW ＞ DW

そうでなければ, CD(Check Data)

CD(Check Data)とは, 見出し語(EW)と定義語(DW)との間の関係が一意的に決らない場合のことである.

2．2 階層関係付けシステムと関係付けデータ

（1）関係付けシステムの概要

[1]語義文から基本的語義文, および, そのタイプと機能表現の表す意味的関係(ρ_{FE})を抽出する.

機能語を含むが機能パターンを含まない語義文などは, チェックデータ(CD1)となる. 機能語と判断された単語が, 定義語の場合があるので, それをチェックするためである.

[2] 基本的語義文から定義語, および, 定義語が複数個かどうか, 定義語に修飾部があるかどうかなどの情報を抽出する.

動植物名が片仮名表記されていたり, 複合語がより適切な定義語とみなせる場合があるので, 片仮名列や漢字列を抽出している.

これら以外の未知語を末尾にもつ基本的語義文は, チェックデータ(CD2)となる.

[3] 2．1の（3）での手順に従い, 定義語と見出し語との関係を決定する.

定義語(DW)と見出し語(EW)との間の関係が一意的に決定できない場合, チェックデータ(CD3)となる.

[4]定義語の意味を区別するために, 概念区別情報, 即ち, ‘読み’（または‘漢字表記’）や‘語義番号’を与える. この処理は, 人間支援である.

語義番号の付加処理において, 語義の区別が一意的に決定できない場合, 未定義記号(アスタリスク：＊)を導入している. これにより複数の上位語が生じることになる. データの検証が必要である.

（2）名詞見出し語についての実験

関係付けシステムを用いて, 新明解国語辞典の名詞見出し語(小見出しも含む.)約57,000語とその語義文約79,000文から, 見出し語と定義語との関係付けデータ約79,000組(一つの語義文から複数の定義語が抽出されている場合があるので)を得ている. この中の約51,500個組に含まれる定義語に概念区別情報が付加されている. 残り, 約27,500組に含まれる定義語には, 語義番号が付加されていない. 内訳は, ‘こと’が約18,900個, ‘もの’が2,300個, 未知語が, 約6,300語であった. 約5,900個の未知語の末尾から部分語が抽出されたが, 約4,600個が適切, 残り約1,300個が不適切とみなされた.

定義語が抽出できなかったか, または, 見出し語との関係が決定できなかったものとして最終的に残ったチェックデータは約2,300個であった.

例, 全体(1012)：…のすべての部分にわたって。

立方根(0010)：Aを…がBである時、Aの称。

図1に, 概念区別情報を付加した関係付けデータの例を示す. これらのデータが以下のシソーラスシステムでの入力用データとなる.

3．関係付けデータに基づいたシソーラスの作成

国語辞典から得られる見出し語と定義語との関係付けデータを利用して, シソーラスを作成するためのシソーラスシステム, およびその実験結果について述べる. ここで, シソーラスが, 次の条件を満たすように, シソーラスシステムを設計している.

①シソーラスでの意味関係は, 関係付けシステムで抽出されるものを用いる.

上位下位関係(＞), 同義関係(≡)

全体部分関係(《), 集合要素関係(∋),

関連関係(R)

②逆関係が登録されている.

③冗長さがない.

即ち, 重複したデータなどがない.

④矛盾がない.

即ち, 異なった関係にあるデータや反対称関係によるループなどが現われない.

基本的語義文の残り	:DW defg	;読み	r	読み EW	defg	FW	FP	α	β	abh	γ
	:顔(0110)	;かお	》	よこっつら【横っ面】	(0110)	横がわ	201	》		000	*
	:海(0010)	;うみ	R	かいてい【海底】	(0010)	底	201	R		000	*
	:気体(0010)	;きたい	<	りゅうたい【流体】	(0010)	総称	203	≡	4	000	
	:液体(001*)	;えきたい	<	りゅうたい【流体】	(0010)	総称	203	≡	4	000	
	:追試験(0010)	;ついしけん	≡	ついし【追試】	(0210)	略	100	≡		000	*
	:警官(0010)	;けいかん	≡	ポリス【ポリス】	(0210)					000	*
相手の	:説(0010)	;せつ	>	こうせつ【高説】	(0010)	敬称	201	≡		000	
日本古来の	:音楽(0010)	;おんがく	>	ほうがく【邦楽】	(0010)					000	
海	:底(0010)	;そこ	>	かいてい【海底】	(0010)	底	201	R		000	
…	:西洋野菜(0010)[2]	やさい;せいようやさい	>	レタス【レタス】	(0010)	一種	201	>		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（‡），
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5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，‘西洋野菜(0010)[2]の2）は，DWの抽出で，一般逆引き辞書とマッチしたキーの長さを示す．

図1　概念区別情報が付加された関係付けデータの例

3．1　シソーラスシステムの概要

シソーラスシステムの構成を図2に示す．シソーラスシステムは，シソーラスデータの作成，シソーラスデータの蓄積，および，シソーラス(データ)の検索の3つのサブシステムからなる．以下，特に断らない限り，単に，作成システム，検索システムなどと呼ぶ．

作成システムは，関係付けシステムで得られたデータ(概念区別情報が付加されている)からシソーラスに必要な情報を抽出し，それをシソーラス用のデータとして作成するためのプログラムである．

蓄積システムは，作成されたデータを検索しやすいデータ構造でファイル化するためのプログラムである．このファイルが第一段階のシソーラスとなる．

検索システムは，シソーラスから見出し語間の階層構造を抽出するためのプログラムである．また，シソーラスの検証にも利用できる．

なお，シソーラスへのデータの追加，削除，修正などを行なうためのシソーラスの更新システムも必要であるが，ここでは言及しない．

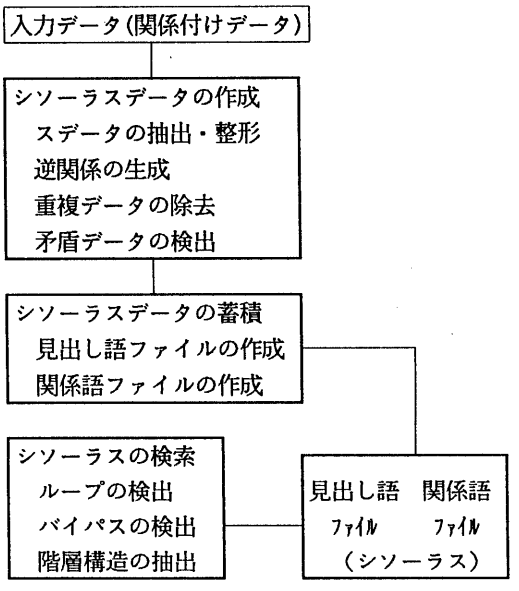

図2　シソーラスシステム

3．2　シソーラスデータ作成システム

入力データとして，階層関係付けシステムで得られたデータ(概念区別情報が付加された)約79,000個と，チェックデータ 2,300個の合計約81,300個を用いる．

（1）シソーラスに必要な情報の抽出

入力データから，次のような情報(関係語データ)を抽出する．

[1]定義語(DW)と関係(ρ)が抽出されていれば，

①EW ρ DW

[2]定義語(DW)が未知語(X)でその部分語(UW)が抽出されていれば，

②EW ρ X，③X＜UW

ここで，ρ＝‘＜’，‘≡’が満たされていれば，

④EW＜UW｛X｝

（2）見出し語と定義語の表記の統一

一般に，辞書の見出し語は送仮名の多様さに対応するために，括弧を用いて表記されている場合が多い．ここでは，括弧を展開して，一番長い表記を標準表記とし，定義語の表記もこれに一致させる．

例：籤引(き) ⇒ 籤引き

（3）逆関係(転置関係)の作成

‘A ρ B’が入力データにあれば，これの逆関係データ‘B ρ^{-1} A’を作成する．

（4）重複データの除去と矛盾データの検出・修正

逆関係の作成などでデータに重複や矛盾が生じることが考えられる．データをソーティングして，次の手順で処理する．

①A≠Bで，A ρ_1 B，A ρ_2 Bの関係があり，

ρ_1とρ_2が同じ → 重複データとして除去

ρ_1とρ_2が異なる → 矛盾データとして検出

②A＝Bで，A ρ_1 Bの関係があり，

ρ_1が同義関係 → 重複データとして除去

ρ_1が同義関係以外 → 矛盾データとして検出

[1]重複データの除去処理の結果

重複データとして，約1,700個を除去した．

[2]矛盾データの検出・修正処理の結果

312個(156対)の矛盾データが検出された．

・同義関係に修正した．(25対)

例　探照燈(0010) ＜ サーチライト(0010)

　　探照燈(0010) ＞ サーチライト(0010)

・上位下位関係を優先した．(129対)

例　大佐(0010) ∈ 佐官(0010)

　　大佐(0010) ＜ 佐官(0010)

・その他のデータはチェックデータとした．

2．3　シソーラスデータの蓄積

（1）ファイルの構成

第一段階のシソーラスである．見出し語ファイルと関係語ファイルとから構成されており，見出し語からその内容(関係語など)が検索できる構造になっている．

見出し語ファイルには，見出し語とポインタ情報が蓄積される．このポインタ情報は，見出し語に対応する関係語ファイルを検索するためのものである．

関係語ファイルには，見出し語に関係(ρ)のある関係語とその概念区別情報などが蓄積される．

（2）シソーラスの項目単位(エントリー)の例

シソーラスの項目単位は，見出し語，関係，関係語から成り，見出し語と関係語には概念区別情報が付加されている．シソーラスのエントリーの例を図3に示す．国語辞典の見出し語にある関係語語には，コロン(：)が付けてあり，未知語には，コロンが付けていない(例‘金属元素’)．

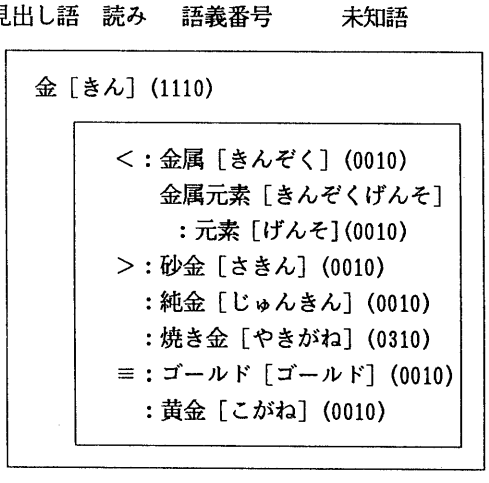

図3　シソーラスの項目単位

3．4　シソーラスデータ検索システム

（1）検索システムと階層構造の抽出例

シソーラスから，単語(概念)間の階層構造を求めるためのプログラムである．即ち，検索すべき単語情報(漢字表記，読み，語義番号)と関係情報に基づき，順次，関係語を求めて行き，見出し語(単語)間の階層構造を検出する．また，ループやバイパスなどの検査にも利用できる．

精根［せいこん］（００１０）
　　＜：体力［たいりょく］（００１０）
　　　　＜：力［ちから］（０３１０）
　　　　　　＜：元気［げんき］（１０１０）
　　　　　　　　＜：気力［きりょく］（００１０）
　　　　　　　　　　＜　精神力［せいしんりょく］　――― 未知語

出力行数＝　　６［行］，検索回数＝　　５［回］，実行時間＝　　１７［ms］

アウトバーン［アウトバーン］（００１０）
　　＜：高速道路［こうそくどうろ］（００１０）
　　　　＜：道路［どうろ］（００１０）
　　　　　　＜：道［みち］（０１１１）
　　　　　　　　＜：所［ところ］（１２１２）
　　　　　　　　　　＜：場所［ばしょ］（０１１０）　――― 一致するものが無い

出力行数＝　　６［行］，検索回数＝　　６［回］，実行時間＝　　２２［ms］

大八車［だいはちぐるま］（００１０）
　　＜：荷車［にぐるま］（００１１）
　　　　＜：車［くるま］（０２１０）
　　　　　　＜：物［もの］（１２１０）
　　　　　　　　＜：物質［ぶっしつ］（０１１０）
　　　　　　　　　　＜　もの［もの］　――― 未知語

出力行数＝　　６［行］，検索回数＝　　５［回］，実行時間＝　　２０［ms］

図４　階層構造の抽出例

抽出された階層構造の例を図４に示す．

このプログラムでは，検索を停止する条件を次のようにしている．

①検索すべき見出し語がシソーラスにない(未知語)場合．
②検索すべき関係語が関係語ファイルにない場合．
③検索すべき関係が関係情報にない場合．
④検索すべき見出し語が，既に同じ経路(パス)で検索されていた場合．（ループ）
⑤検索すべき見出し語が，既に別の経路(パス)で検索されていた場合．（クローズパス）
⑥予め設定していた，各種の制限値を越えた場合．

４．国語辞典に基づくシソーラスについての考察

４．１　シソーラスの概要

（１）シソーラスエントリーの調査

シソーラスの見出し語の数	約79,000個
関係語の総数	約140,000個
見出し語当りの関係語の数	約1.77個
見出し語当りの関係の数	約1.19個
上位語を持たない語の数	約 12,500個
下位語を持つ語(極大語)の数	約1,500個
下位語を持たない語(孤立語)の数	約11,000個
上位語を持つ語の数	約66,500個
下位語を持つ語(中間語)の数	約5,500個
下位語を持たない語(極小語)の数	約61,000個

（２）階層構造の深さ

下位語を持たない語(孤立語・極小語)を始点に，上位－下位関係での階層の深さ(階層のレベル)を調べた．平均の深さは約 3.2である．

（３）階層関係のループ

ある単語(見出し語)Wから出発して，ある関係の基で，関係語を求め，その関係語を見出し語として，同じ関係にある関係語を求め，…と，順次，見出し語を求めて行く過程で，再び，自分自身Wにもどる経路が出来るとき，ループと呼ぶ．反対象関係でループが存在すると，関係に矛盾が生じることになる．ただし，この場合，そのループを構成している単語が，互いに同義語(類義語)とみなせれば同義関係に

修正することにより矛盾の解消は可能である．

今回，下位－上位関係(<)でのループは，検出できなかった．同義関係を考慮したループについて調査を進めている．

（4）階層構造でのバイパス

ある関係の基で，ある単語を始点にした複数のパスが，共通の語に到達する場合に，始点の単語から直接その共通の語へ到達するパスがあるとき，そのパスをバイパスと呼ぶ．バイパスは，推移律が成り立つ関係では，一般に，冗長なパスと考えられる．しかし，後で述べるように，複数上位語での複数パスの中で優先的なパスの選択や，一種の推論の効率化などに有用になると思われるので，特に，重複など明らかに冗長と考えられる場合以外は，削除しないようにしている．

今回作成したシソーラスでは，約200個のバイパスが検出された．以下に例を示す．

例：太陽[たいよう](0010)
　　<:天体[てんたい](0010)
　　<:恒星[こうせい](0010)
　　<:星[ほし](0110)
　　<:天体[てんたい](0010)

例：五目ずし[ごもくずし](0010)
　　<:寿司[すし](2010)
　　< 散らし寿司[ちらしずし]
　　<:寿司[すし](2010)

4．2　極大語・孤立語の処理

上位語を持たない見出し語は，階層構造上でそれ以上，上位関係で連結できなくなる．このような語は，約12,500個あった．この中で，下位語を持つものを極大語と呼び，下位語を持たないものを孤立語と呼んでいる．ただし，語義番号の違いにより表記は同じでも異なる語として計数されている．他の場合も同様である．なお，上位語を持つ語の中で，下位語を持つものを中間語，下位語をもたないものを極小語と呼んでいる．

（1）極大語の中で，‘もの’，‘こと’，および，未知語となる複合語については，文献(4)(6)で考察しているので割愛する．

（2）未知語でない見出し語で，複合語と見做されるものは，一般に，末尾の語がその複合語の上位語である場合が多い．従って，語構成を考慮して上位語の抽出が可能となる．

例：圧搾空気<空気

（3）同義語を持つ場合，その語の関係語を調べ，その中に上位語があれば，間接的に，上位語が得られることになる．上位語がない場合は，さらに同義語があれば，同様な処理を行う．この様にして，間接的であるが，極大語でなく中間語にできると考えられる．

例：イースト[イースト](0010)
　　＞：パン種[パンだね](0010)
　　≡：酵母[こうぼ](0010)
　：酵母[こうぼ](0010)
　　<：菌類[きんるい](0010)
　　：植物[しょくぶつ](0010)
　　≡：イースト[イースト](0010)

（4）同義語を利用した同様な考えで，極小語についても，間接的に中間語にできる場合がある．

例：ＣＭ[シーエム](0010)
　　≡：コマーシャル[コマーシャル](0010)
　：コマーシャル[コマーシャル](0010)
　　<：放送[ほうそう](0210)
　　≡：ＣＭ[シーエム](0010)

4．3　複数上位語と複数パス

（1）オープンパスとクローズドパス

一つの語義文から，複数の上位語が抽出される場合があり，階層構造上で複数パスが生じる．複数パスは，クローズドパスとオープンパスとに分けられる．オープンパスとは，複数のパスが共通の上位語(上位概念)に到達しない場合であり，クローズドパスとは，複数のパスが共通の上位語に到達する場合である．オープンパスは，孤立した極大語が多数存在する場合に生じる可能性がある．また，バイパスは，クローズドパスの特殊な場合である．

オープンパスについて，現在調査中であり，極大語の処理や複数継承の問題とも関連してくると思われる．

以下，バイパスの利用について，少し検討する．

（2）パスの優先的選択

例えば，下図の例で，‘教養(0211)’から‘学問(0210)’を介して上位語へのパスを求める場合，‘知識(0112)’へのバイパスがあれば，

‘教養(0211)’≒‘学問(0210)’≒‘知識(0112)’

のパスを優先的に選択にしようというわけである．従って，そのような検索が行われたときは，パスが決ればバイパスは潜在的なものとして取り扱うこと

にすればよい.

複数パスでの上位語選択の発散を防ぐのに役立つと思われるし，未知語の場合に，意味の分化のモデル化への手掛りとなるのではないかと思われる.

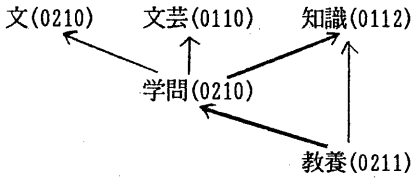

図5　パスの選択へのバイパスの利用の例

（2）推論での利用

例えば，下図の例で，下位語‘ポインター’から上位語を求める場合，‘猟犬’か‘番犬’かどちらかのパスを選択する必要がでてくる．この場合，情報がないときは，取り敢えず，複数パスを無視して‘ポインター’から‘犬’への一時的なバイパスを設定して，後で必要な情報が得られたらこのパスを解消しようというわけである.

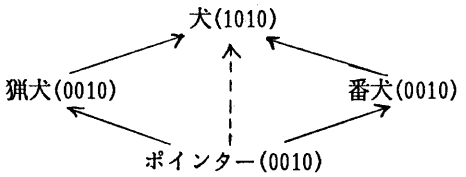

図6　一時的なバイパス設定の例

5．おわりに

新明解国語辞典の名詞見出し語約57,000語(小見出しを含む)の語義文約79,000文から，階層関係付けシステムを用いて得られた出力データに基づいて，見出し語データ約80,000語をもつシソーラスを試作し，その結果について考察した.

情報の欠落などによる，不備は多いが，更新システムを作成し，改良を進めてゆく予定である.

なお，本稿で述べた実験システムは，長崎大学情報処理センターFACOM-M760上でPL/1で実現されているが，現在，ＰＣまたは，ワークステーション上で利用できるようにＣ言語で移植中である．データの転送などにＮＴＴから借用中のＥＬＩＳ-Ｉを利用している.

今後の主な課題として次のようなものがある.

（1）単語の多義，概念の揺れ，曖昧さなどが単語間の階層関係を求める場合の大きな問題となっている．本稿では，単語間の関係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.

（2）階層構造上での単語間の跳躍と，特定の単語への下位語の集中が問題となっている．極大語の階層構造への統合，下位語の分類などを行ない，中間の語の補間を行なう必要がある．同義関係の積極的な利用や，語義番号の統合化などを検討している.

（3）単語間の階層関係を基礎に，単語の意味記述を行なうためには，語義文中の修飾部の解析を始め，補足的説明や用例などの解析も必要である.

（4）複数パスによる冗長性の問題や複数継承の問題などがある．複数パスでのパスの優先的選択や推論の効率化なども興味ある課題である.

謝辞　資料の収集・整理，プログラムの開発などに松崎功君(現在：ＮＢＣ)，兵頭竜二君(現在：富士通)，井上淳君(現在：キャノン)を始め，研究室の卒論生諸氏の協力を得た．ここに，感謝の意を表す.

なお，本研究の一部は，文部省科学研究費特定研究，ＥＤＲ奨学寄付金およびＡＴＲ受託研究費による.

参　考　文　献

(1) 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) 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 (1986)

(3)鶴丸，日高，吉田：単語間の上位－下位関係の自動抽出，情処学研報(FI)，Vol.86，No.3，pp.1-8 (1986)

(4) 鶴丸，兵頭，松崎，日高，吉田：語義を考慮した単語間の階層構造の抽出について，情処研報(NL)，Vol.87,No.84,pp.9-16(1987)

(5) 徳永，奥村，田中：概念階層への視点の導入，情処学論，Vol.30，No.8，pp.970-975(1989)

(6) 鶴丸，日高，吉田：シソーラス構築への国語辞典の応用に関する一考察，信学技報(NLC90-10)，Vol.90，No.116，pp.33-40 (1990)