

国語辞典からの情報抽出とその構造化

鶴丸弘昭・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

自然言語の意味処理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。しかし、膨大な単語について、その意味情報を何からどのような方法で収集するか、また、どのような構造で表現するかという基本的ではあるが、難問があり、その実現に困難をきたしている。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。⁽⁴⁾⁽⁵⁾⁽⁶⁾⁽⁷⁾⁽⁸⁾ この研究は、次の3つの部分に分けられる。

1. 国語辞典の記述内容を計算機で利用可能にするための種々の計算機プログラムを開発する。(活用システムの開発: 情報の抽出と構造化)
2. 国語辞典の意味記述文(語釈義文)から、見出し語(EW: Entry Word)と(階層)関係のある語(DW: Definition Word)を抽出し、階層関係が成立するEWとDWとを利用して、単語間の階層構造を求める。(単語間の階層関係付けシステムの開発: シソーラスの自動作成に関する試み)

3. 国語辞典の記述内容から得られる情報を利用して、単語間の階層関係を基礎に、フレーム(frame)構造で単語の意味を定義する⁽²⁾⁽¹²⁾⁽¹³⁾(意味辞書の開発)これらはさらに細かいテーマを含んでいる。

3. に関する概要については、COLING82⁽⁹⁾で発表した。本報告では、1., 2. に関して、情報の抽出を中心に、考察する。

図1に国語辞典からの情報抽出処理についての手順を示す。以下、図1の手順に従ってDWの抽出までの処理について説明する。

ここで、我々が用いた辞書は、三省堂出版の「新明解国語辞典」(金田一京助等編)で、親見出し約6万語、子見出し約8,000語を持っている。この辞典は電子技術総合研究所推論機構研究室と三省堂との共同で磁気テープ化されている。⁽²⁾ 我々は、三省堂との契約によりこの磁気テープを利用している。

2. 国語辞典からの情報の抽出

2. 1 国語辞典データの記述形式と抽出項目

磁気テープ化された国語辞典から見出し語や正書法などの見出し部を取り出し、意味部(見出し語を除いた残りの部分)と分離してファイル化するプログラムが、すでに開発されている。⁽²⁾⁽⁶⁾ しかし、国語辞典を高度に活用するには、この意味部に含まれている語釈義文、用例、補足的説明、反対語、参照語などを容易に利用できるようにしておかなければならない。

図2に、今回、抽出の対象とした情報(項目)間の関係を構造化して示す。

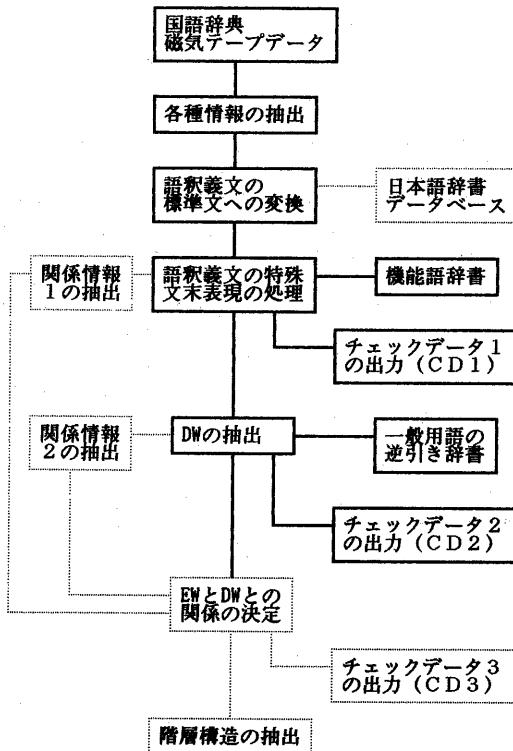


図1. 国語辞典からの情報抽出の手順

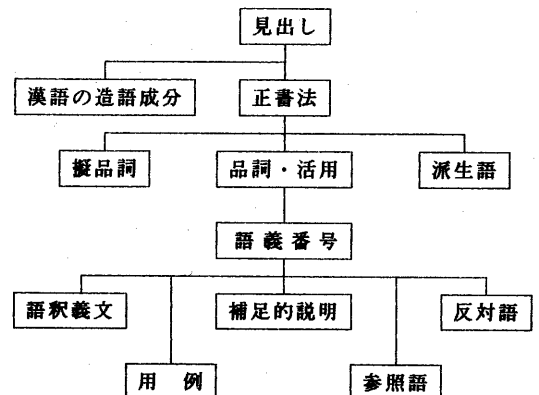


図2. 抽出の対象にした項目間の関係

2. 2 抽出処理の概要

意味部の特徴として、

- (1) 見出し（正書法）の直接の意味内容（意味記述部）の他に、同音語や子見出しの意味内容が含まれている。
- (2) 各項目の出現順序は、必ずしも決まっているわけではない。例えば、語義番号の前に品詞が現われる場合もあるし、正書法の前に語義番号（㊦㊦…）が現われる場合もある。
- (3) 全ての見出しが、意味部に正書法（漢字表記）や語義番号を持っているわけではない。
- (4) 各種の情報（項目）を説明するために、特別な記号（⇒、←、㊦など）、略号（（略）、〔雅〕、〔名〕など）、略語（一する、イ、（五）、（自）など）が用いられている（以下、特殊記号と呼ぶ）。

などがある。

我々は、複雑な構造を持つ意味部から、与えられた見出し（正書法）の意味記述部の情報（図2に示した項目）を抽出するために、情報（項目）が示す範囲を段階的に絞り込んでいく方法を取った。その手順の概要は次の通りである。

- (1) 正書法（または子見出し、同音語）の直接の意味内容を取り出す（意味記述部の抽出）。
- (2) 語義番号（㊦㊦…）が示す範囲を取り出す。
- (3) 語義番号（㊦㊦…）が示す範囲を取り出す。
- (4) 特殊記号を手掛りに各項目に対応するデータを抽出する。

ところで、1つ（1組）の特殊記号が、必ずしも1つの内容を示しているわけではない。この様なあいまいさを除去するためには、他の記号や文字との相互関係にも着目する必要がある。簡単な例として、かぎ括弧（「『、』」）は用例を示す他に、語釈義文や補足的説明（「〔、〕」）の中にも現われる。

この場合の処理としては、①〔…〕の中に現われる「…」は、補足的説明中のかぎ括弧とみなし、用例として抽出しないようにする。②終わるかぎ括弧（「』」）の次に現われる記号または文字を検査することにより、用例か語釈義文の一部かを判定する。

図3に、国語辞典磁気テープデータの見出し（正書法）とその意味部の例を示す。

図4に抽出処理プログラムでの実行例を示す。

図4の例は、図3の見出し（正書法）が与えられた場合、その意味部から情報（項目）を抽出したものである。ここで、プリンター文字の都合により、特殊記号の一部が国語辞典中での記号と異なっている（語義番号を表わす「㊦」、「㊦」、「㊦」、「㊦」が各々「1」、「2」、「(1)」、「(2)」に、アクセントを示す「@」が「O」に替っている）。

意味記述部の中で、補足的説明や用例がどの語釈義

しゅうかい【周回】

語義* - 1	派生形	一する
	補足	〔何かの回りを〕
	補足	〔回る回数をも指す〕
	用例	「地球を一する人工衛星・一飛行」
	釈義文	@回ること。@#
語義* - 2	用例	「一三十キロの自然公園」
	釈義文	「周回」の意の漢語的表現。#

すすぐ【漱ぐ】

品詞	(他五)
語義2	
派生語	
可能	すすげる
補足	〔1.(2)は、雪ぐとも書く〕
釈義文	口をゆすぐ。うがいをする。@

図4. 抽出処理の実行例

文についてのものか、は重要な情報である。また、それらが語釈義文の前にあるか、後にあるかなども大切な情報である。このために、図4に示すように「@」（補足的説明の位置）、「#」（用例の位置）が自動的に挿入される。

2. 3 考察と今後の課題

我々は、情報抽出実験のために、磁気テープ化された国語辞典から、実験用のデータとして1/20縮小データを作っている⁽⁴⁾⁽⁵⁾ 1/20縮小データには、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を利用した。また、語釈義文の構造的特徴や特殊な記述の調査には、1/20縮小データに限定せず、随時、通常の国語辞典⁽¹⁾や磁気テープデータを利用した。

この1/20縮小データの見出し語 3,206語について、抽出実験を行なった結果、正しく抽出されたもの 3,180語（99.2%）、データエラーとして検出されたもの 20語（0.6%）、誤って抽出されたもの 6語（0.2%）であった。ここでのデータエラーは、意味部に含まれている左括弧と右括弧の対を調べることによって検出される。

抽出誤りの内訳は、

①正書法の意味記述部抽出の失敗（4語）

②特殊記号の同定の失敗（2語）

であった。これらについては、正しい抽出ができるようにプログラムを修正している。

しゅうかい【周回】

20シウクワイ【周回】(1)一する〔何かの回りを〕回ること。〔回る回数をも指す〕「地球を一する人工衛星・一飛行@」(2)「周回」の意の漢語的表現。「一三十キロの自然公園」

すすぐ【漱ぐ】

30(他五) 1.「濯ぐ」(1)「水の中にたっぷりつけて」よごれを洗い落とす。「洗濯物センタクモノを一」(2)「りっぱな行いをして」恥・不名誉などのつぐないをする。そそぐ。「汚名を一」2.「漱ぐ」口をゆすぐ。うがいをする。(可能)すすげる○〔1.(2)は、雪ぐとも書く〕

図3. 見出し（正書法）とその意味部の例

情報抽出での問題点として、

- (1)漢字表記の正書法がない場合や、正書法と(大)語義番号とが複雑な組み合わせになっている場合がある。人間にはすぐ理解できるが、抽出範囲を機械的に決定することが、困難となる。
- (2)特殊記号にあいまいさがある。
- (3)データエラーがあるので、その対策を考えておく必要がある。1/20縮小データに関しては、情報抽出で重要となる特殊記号については修正している。括弧については、対を調べることでエラー検出を行っている。記号や文字のエラーは随時、訂正できるようにしている。
- (4)特殊な記述形式がある。
 - a.用例の後に語釈義文が来ている。
 - b.小語義番号で示される範囲内にⒶⒷが使用されている。
 - c.派生語の記述があいまい(語釈義文との区別が困難)なものがある。
 - d.一つの意味部の中で、異なった語義番号に対して、同じ漢字表記の正書法が用いられている。

などがある。

今後の研究課題として、次のようなものがある。

- (1)補足的説明中の内容の分析(「『…』」の中の処理)
 - a.語原情報の抽出 [←…]
 - b.参照語情報の抽出 [⇒…]
 - c.位相情報の抽出
 - d.同義的意味情報の抽出(説明文の分析が必要)
- (2)用例の分析(「『…』」の中の処理)
 - a.個々の用例の抽出
 - b.省略部分の復元
 - c.用例中の語句の説明の抽出 「… [= …] …」
- (3)子見出しの利用
 - a.子見出しに「ひらがな見出し」をつける
 - b.子見出しの中の漢字の読みを削除する
 - c.品詞を付加する
- (4)国語辞典のデータベース化
使用目的に応じて、各情報間の相互参照が決まると思われる。現在、図2に示した情報(項目)間の関係を基本にしてデータベース化を考えている。

3. 語釈義文とその標準文への変換

3.1 語釈義文の特徴

語釈義文(DS: Definition Sentence)は、見出し語(EW)の意味・用法を言葉で説明したものであり、次のような特徴を挙げることができる。

- (1)DSの中にEWと階層関係にある語(DW)が含まれている場合が多い。
- (2)DSでは、DWがいくつかの側面から規定されている。
- (3)DWは文末側に現われることが多い。
- (4)DSの文末側に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。この特殊文末表現には、EWの語用・語源に関する情報を含んでいるものがある。
- (5)DSは、記述を簡略にするため、括弧やドット(・)を使った省略表現が多い。
- (6)漢字に読み(ルビ)が添えてある場合がある。

(7)DS中に現われる動植物名はカタカナ表記になっている。

(8)DS中で用いられる単語は必ずしも漢字表記の正書法になっていない場合(ひらがな書きなど)がある。

(9)DWの抽出という立場から見ると、DSの記述方法に統一性がないと思われる場合がある(人間の使用を前提にしているため)。

(10)DWが「こと」や「もの」になっている場合が多い。

(8)(9)は機械処理を困難にする要因の一つになっている。

3.2 語釈義文の標準文への変換

語釈義文には、3.1の(5)(6)に挙げたように、括弧やドット(・)による省略、漢字の読みの付加がある。語釈義文を機械処理するためには、語釈義文を通常の日本語文表現に変換しておく都合がよい。特に、一般の日本語文の機械処理用に開発されたプログラムを語釈義文の処理に利用するためには、必要である。このための変換処理を標準文変換と呼ぶ。標準文変換は、(1)漢字の読みの除去と、(2)括弧とドットの処理(括弧の削除)からなる。

(1)漢字の読みの除去

国語辞典の中では漢字の読みはルビになっているが、磁気テープデータでは、丸括弧「()」の中にカタカナで表現されている。このため、省略表現と同じ形式になるので、まず最初に語釈義文の中から漢字の読みに相当する部分を見つけ出し、除去しておく必要がある。今回は、第一近似として、漢字列の直後に丸括弧が来て、その括弧の中がカタカナ列であれば、その漢字列の読みとみなして、除去するようにした。約4,150個の文に当たってみたところ、今回の処理で誤った場合として、

a.丸括弧の中が外来語である。

例。ようが【陽画】:…、写真・原版(フィルム)。

b.丸括弧の中に漢字とその読みの両方がある。

例。にのとり【二の酉】:…の酉の日(の市イチ)がであった。

漢字の読みは語釈義文の「…」の中(丸括弧なしでカタカナ表記)や、小見出しの中(丸括弧なしでひらがな表記)にも現われる。また、逆に語釈義文中の記述がひらがな書きになっていて、その漢字表記の正書法が丸括弧の中に記述されている場合がある。

これらの漢字の読みの除去を本格的に行なうために、「日本語単語辞書」⁽¹⁵⁾(九州大学大型計算機センター公用データベース)の利用を計画している。

(2)括弧とドットの処理(括弧の削除)

語釈義文中に用いられている簡略表現には次の3つの型がある。

i. 省略(可能)型

…αα…α(ββ…β)γγ…γ,(.)

標準文 { …αα…αββ…βγγ…γ,(.)
 { …αα…αγγ…γ,(.)

ii. 置き換え型(a)

…αα…α・ββ…β(γγ…γ)δδ…δ,(.)

標準文 { …αα…αββ…βδδ…δ,(.)
 { …αα…αγγ…γδδ…δ,(.)

iii. 置き換え型(b)

$\dots \alpha_1 \alpha_2 \dots \alpha_i \cdot (\beta \beta \dots \beta) \gamma \gamma \dots \gamma, (.)$
 標準文 $\left\{ \begin{array}{l} \dots \alpha_1 \alpha_2 \dots \alpha_i \gamma \gamma \dots \gamma, (.) \\ \dots \beta \beta \dots \beta \gamma \gamma \dots \gamma, (.) \end{array} \right.$

置き換え型(b)の変換処理で問題になるのは、 $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_i$ の範囲を決める(すなわち、 α_i の位置を見つけ出す)ことである。今回の調査では、置き換え型(b)が語釈義文に現われる場合は、 α_i が文頭になっているか、または、'、'の直後に来ているかのいずれかであった。

図5に、(2)の処理の実行例を示す。

あたり【当(た)り】：さわった・(接触した)時の感じ。

標準文 $\left\{ \begin{array}{l} 1. \text{さわった時の感じ。} \\ 2. \text{接触した時の感じ。} \end{array} \right.$

とうせい【騰勢】：(物価が)上がる・勢い(傾向)。

標準文 $\left\{ \begin{array}{l} 1. \text{上がる勢い。} \\ 2. \text{上がる傾向。} \\ 3. \text{物価が上がる勢い。} \\ 4. \text{物価が上がる傾向。} \end{array} \right.$

図5. 標準文変換処理の実行例

変換処理の実験の結果、失敗例に次のようなものがあった。

a. 語釈義文中の記述がひらがな書きになっていて、その漢字表記の正書法が丸括弧の中に記述されている。
例. かしゃく【仮借】：かしゃ(仮借)。

b. 人間に理解可能な表現である。

例. おお【大】：「大い(な)・大変(な)・たくさん」の意を表す。

はっかん【発刊】：新しい雑誌・新聞などを(初めて)出版すること。

c. 外来語の原語が丸括弧の中に来ている。

例. しゃおう【沙翁】：「シェークスピア(Shakespeare)」の漢語的表現。

4. 語釈義文の特殊文末表現の処理

4.1 特殊文末表現の種類と機能

3.1で述べたように、一般に見出し語(EW)と階層関係にある語(DW)は、語釈義文(DS)の文末に現われる場合が多い。しかし、必ずしもそうとは限らず、DSの文末に特殊文末表現(DWとEWとの関係を規定する表現)が現われる場合がある。DSからDWを抽出し、EWとDWとの関係付けを行うためには、この特殊文末表現の構造、種類、および機能を明らかにしておく必要がある。我々は、磁気テープデータから約3,500語の名詞見出し語を抽出し、その語釈義文(一つの見出し語が複数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語釈義文を抽出している)の第一文について調査を行なった。調査のために、抽出した語釈義文を標準文に変換し、文末側から(KWIC)ソーティングした資料(約7,000文)を作成した。

表1. 語釈義文の機能パターンの型と頻度

型	機能パターン(FP)	頻度
100	「...DW」+ (など) + σ^* + FE.	1020
200	...	...
201	...DW+ (など) + の + FE.	342
202	...DW+ (など) + W_{202} + の + FE.	9
203	...DW+ (など) + と + DW+ (など) + との + FE.	8
300	...	...
301	...DW+ (など) + を + σ^* + W_{301} + FE.	28
302	...DW+ (など) + に対する + FE.	1
400	...DW+ など.	27

ここで、 σ^* ：任意の文字列
 W_{202} ：意、の意、として、,、の中、の種、
 W_{301} ：言、の、いう、言った、呼ぶ、呼んだ、
 指す、指した

また、1/20縮小データの中の名詞見出し語約1,500語についても同様な資料(約3,000文)を作成し、調査を行なった。

調査の結果、計算機処理を考慮して特殊文末表現を(文末)機能パターン(FP: Functional Pattern)と(文末)機能語(FE: Functional Expression)とに分けた。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。FPとして約40種類が調べられたが、本稿ではそれらを表1に示す型に統合した。型の特徴は次の通りである。

100型: DSに「」が含まれている。

200型: FEの直前の文字が「の」である。

300型: FEが連体修飾されている。

400型: DSの文末が「など」で終わっている。

FPの約71%が「」を含んでいる。この場合、ほとんどのDWは「」の中に現われる。代表的な機能パターンの型は100型と201型で、全体の約95%を占めている。

表2は、FEの種類、FPとの組合せによるDWとEWとの関係、および頻度を示したものである。

現在、139個のFEを求めている。表2は、FEに含まれている文末語(文字)の頻度が高い順に整理されている。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(W))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。

FEには、DWとEWとがどのような位相(見方)での階層関係(主に同義関係)かを明らかにするものがある。これは、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。

FEを含むDS(第一文について)は全体の約16%である。その中でFPを満足するものは約88%(全体の約14%)である。残りの約12%(全体の約2%)については人間がチェックするように考えている。全体の約84%に対しては、文末語がDWであると思われる。

4.2 機能語辞書

機能語辞書(FED)は、4.1で求めた機能語(FE)を拡張B-treeの構造⁽⁶⁾でファイル化したものである。

表2. 機能語の種類、FPとの組合せによる
DWとEWとの関係およびその頻度

文末語	機能語 (FE)	型 (FP)	頻度	関係 (DW:EW)
表現	漢語的表現	100	357	=>
	雅語的表現	100	178	=>
	古語的表現	100	35	=>
語	老人語	100	73	=>
	丁寧語	100 201	33 1	=> =
	女性語	100	33	=>
	異称	100 201	24 20	=> =>
	敬称	100 201	5 20	= =>
称	総称	201 203	44 2	=> <
	称	100 201 202	5 29 1	=> => <
言い方	言い方	301	4	=>
	古風な言い方	100	1	=
	新しい言い方	100	1	=
一つ	一つ	201 202 203	28 1 1	> > >
	一つ一つ	201	1	W
言葉		301	10	=R
略	略	101 201	64 7	=< =<

文末側からマッチングを取るため逆引き用になっている。

FEDは、次の内容を持っている。

- (1) 逆向列の機能語 (FE)
- (2) FEの文末側の単語の文字数
- (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によって規定するDWとEWとの関係情報

4. 3 Skeleton Sentence の抽出

Skeleton Sentence (SS) とは、語釈義文 (DS) が特殊文末表現を含まない場合は、DSそれ自体であり、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。すなわち、SSでは、文末にDWが現われる。

SSの抽出手順は次の通りである。

Step 1. (機能語 (FE) の判定)

入力文 (or DS) の文末表現と機能語辞書 (FED) のFEとのマッチングを最長一致法で行なう。マッチングしなければ、Step 5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、文末側のFPの型を決定する。FP判定処理ルーチンは、4. 1の表1をもとに作成している。満足するFPがない場合、FPがあった場合でもそのFPの関係情報が '?' である場合、または複数個のFE ('など' を除く) がある場合は、Step 6へ。

Step 3. (関係情報 (1) の蓄積)

Step 1でマッチングしたFEが持っている関係情報をスタックする。 'など' がDSに含まれていたかどうか、FPの型は何かも蓄積する。

Step 4. (FPの削除)

入力されたDSよりFPを削除し残りの部分を入力文として、Step 1へ。 100型の場合は、 'J' の次の2文字を入力する ('など' の判定のため)。

Step 5. (Skeleton Sentence (SS) の抽出)

入力されたDSが 100型のFPを持つ場合は、 'I' と 'J' で挟まれた部分の語句がSSとなる。入力文 (or DS) がFEを含まない場合は、入力文がそのままSSとなる。ここで、抽出されたSSの文末語が '時' の場合は、Step 6へ。

Step 6. (チェックデータ 1 (CD1) の出力)

DSがFEを含み、FPを満足しない場合や、複数個のFEを含む場合などを、検出する。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。

実験に用いた語釈義文の総数は 2,824文である。

1/20縮小データの中の名詞見出し語の語釈義文の第一文を対象にした。

抽出実験の結果、SSが抽出された語釈義文の数は 2,725文である。そのうち、語釈義文にFEが含まれていないものが 2,404文で全体の約85%であり、FPを含んでいるものが 321文で約11%である。チェックデータ 1 (CD1) として出力された語釈義文は99文で約 4%である。

(1) SSが抽出されたDSの中で、FPを含んでいるもの

(321文中) の内訳

100型	213文	201型	98文
200型	99文	203型	1文
300型	3文	301型	2文
400型	6文	302型	1文

(2) チェックデータ 1 (CD1) として出力された語釈義文 (99文) の内訳

- FEがありFPを満足しないもの 85文
- FEがありFPを満足するもので、
 - FEの関係情報が '?' である 10文
 - 抽出された文の文末が '時' である 2文

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)⁽²⁰⁾の抽出と調査。いくつかの文について語釈義文の形態素解析⁽¹⁸⁾を試みたが、データエラー等の影響であり良い結果は得られなかった。データエラーが訂正された磁気テープデータを利用する必要がある。

(4) 語釈義文の、格情報を用いた係り受け解析⁽¹⁷⁾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、意味辞書の開発では、動詞、形容詞、副詞等の意味の定義⁽¹¹⁾⁽¹³⁾も必要である。

なお、本稿で報告した抽出処理実験に、九州大学大型計算機センターのFACOM M-382、長崎大学情報処理センターのFACOM M-180 II AD を利用した。

最後に、実験システムの計算機処理を手伝っていた長崎大学工学部修士1年水野浩司君、ならびに、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、吉村賢治助手(現在、福岡大学工学部講師)に感謝の意を表します。

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 (1977)
- (3) 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(昭54.7)
- (4) 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 (1978)
- (5) M. Nagao, J. Tuijii, Y. Ueda, M. Taiyama: AN ATTEMPT TO COMPUTERIZED DICTIONARY DATA BASES, Proc. COLING80, PP.534-542 (1980)
- (6) 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(昭54.2)
- (7) 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年度科学研究費補助金試験研究(1) 研究成果報告書(昭57.2)
- (8) 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)
- (9) A. Michiels, J. Mullenders, J. Noel: EXPLOITING A LARGE DATA BASE BY LONGMAN, Proc. COLING80, PP.374-382 (1980)
- (10) A. Michiels, J. Noel: APPROCHES TO THESAURUS PRODUCTION, Proc. COLING82, PP.227-232 (1982)
- (11) 栗原俊彦、吉田将、鶴丸弘昭、藤田毅：言語と思考のシミュレーション、学習研究社、情報社会科学講座第4巻(昭52)
- (12) 吉田将：自然言語理解(知能情報処理とロボット特集)、電子通信学会誌、Vol.65, No.4 (1982)

- (13) S. Yoshida: CONCEPTUAL TAXONOMY FOR NATURAL LANGUAGE PROCESSING, in Kitagawa, T.(ed), Computer Science & Technologies, Ohmsha, Japan (1982)
- (14) S. Yoshida, H. Tsurumaru, T. Hitaka: MAN-ASSISTED MACHINE CONSTRUCTION OF A SEMANTIC DICTIONARY FOR NATURAL LANGUAGE PROCESSING, Proc. COLING82, PP.419-424 (1982)
- (15) 吉田将、日高達、稲永紘之、田中武美、吉村賢治：公用データベース日本語単語辞書の使用について、九州大学大型計算機センター広報、Vol.16, No.4 (昭58.7)
- (16) 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 (昭57.10)
- (17) 日高達、吉田将：格文法による日本語の構文解析、自然言語処理技術シンポジウム論文集、PP.41-PP.46 (昭58.6)
- (18) 吉村賢治、日高達、吉田将：文節数最小法を用いたべた書き日本語文の形態素解析、情報処理学会論文誌、第24巻、第1号(1983)
- (19) 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)
- (20) 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)
- (21) 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)
- (22) 鶴丸弘昭、内田彰、日高達、吉田将：意味辞書開発のための国語辞典の活用について、情報処理学会第28回(昭和59年度前期)全国大会、3M-5 (1984)
- (23) 鶴丸弘昭、内田彰、日高達、吉田将：国語辞典を利用した単語間の階層付けシステムの開発について、情報処理学会第28回(昭和59年度前期)全国大会、3M-6 (1984)