

用例や表層表現を用いた日本語文章中の  
指示詞・代名詞・ゼロ代名詞の指示対象の推定

村田真樹 長尾真

京都大学工学部 電気工学第二教室

〒606 京都市左京区吉田本町

TEL:075-753-4985 {mmurata,nagao}@kuee.kyoto-u.ac.jp

あらまし

日本語文章における代名詞などの代用表現の指す対象が何であるかを把握することは、対話システムや高品質の機械翻訳システムを実現するために必要である。そこで、本研究では用例、表層表現、主題・焦点などの情報を用いて指示詞・代名詞・ゼロ代名詞などの指示対象を推定する。従来の研究では、代名詞などの指示対象の推定の際には意味的制約として意味素性が用いられてきたが、本研究では対照実験を通じて用例を意味素性と同様に用いることができることを示す。また、連体詞形態指示詞の推定には意味的制約として「AのB」の用例を用いることを提案する。

キーワード 機械翻訳, 照応, 指示対象, 用例, 表層表現, 対照実験

An Estimate of Referents of Pronouns in Japanese Sentences  
using Examples and Surface Expressions

Masaki Murata Makoto Nagao

Department of Electrical Engineering, Kyoto University

Yoshida-honmachi, Sakyo, Kyoto 606, Japan

TEL:075-753-4985 {mmurata,nagao}@kuee.kyoto-u.ac.jp

Abstract

It is necessary to clarify referents of pronouns in machine translation and conversational processing. We present a method of estimating referents of pronouns in Japanese sentences using examples, surface expressions, topics and focuses. In conventional work semantic markers have been used for semantic constraints. On the other hand, we use examples for semantic constraints and show that examples are as useful as semantic markers through control experiments. We also propose to use examples of the form "A of B" for estimating referents of demonstrative adjectives.

key words machine translation, anaphora, referent, example, surface expression, control experiment

## 1 はじめに

日本語文章における代名詞などの代用表現を含む名詞の指す対象が何であるかを把握することは、対話システムや高品質の機械翻訳システムを実現するために必要である。そこで、我々は用例、表層表現、主題・焦点などの情報を用いて名詞の指示対象を推定する研究を行った。普通の名詞の指示対象の推定方法はすでに文献<sup>(1)</sup>で述べた。本稿では指示詞・代名詞・ゼロ代名詞の指示対象の推定方法について説明する。

代名詞などの指示対象を推定する研究として過去にさまざまなものがあるが<sup>(2)(3)(4)(5)(6)</sup>、これらの研究に対して本研究の新鮮さは主に次のようなものである。

- 従来の研究では代名詞などの指示対象の推定の際に意味的制約として意味素性が用いられてきたが、本研究では対照実験を通じて用例を意味素性と同様に用いることができることを示す。
- 連体詞形態指示詞の推定には意味的制約として「A の B」の用例を用いる。
- 「この」が代行指示になりにくいという性質を利用して解析を行なう。
- 指示詞による後方照応を扱っている。
- 物語文中の会話文章の話し手と聞き手を推定することで、その会話文章中の代名詞の指示対象を推定する。

## 2 指示詞の指示対象の推定方法

指示詞はいろいろな形で出現するので、それぞれに対して詳細に規則を作ることによって高精度に解析することができる。

### 2.1 名詞形態指示詞の指示対象の推定方法

名詞形態指示詞については名詞形態指示詞で人を指すことがないという性質を用いることによって指示対象の候補をかなり削減することができる(判定規則 1, 2)<sup>1</sup>。例えば、次の例文中の「それ」の指示対象は「コンピューター」であるが、「それ」の近くにある「コンピューター」以外の名詞は人を表すものしかないので、指示対象が「コンピューター」であることがわかる。

太郎は最新のコンピューターを買いました。  
ジョンに早速それを見せました。(1)

「これ」「それ」などの指示詞は名詞のみを指示するとは限らず、事態を指示する場合がある。

天狗達は問もなくやってきて、前の晩のように歌ったり踊ったりし始めました。  
おじいさんはそれを見て、こんな風に歌い始めました。(2)

この例の「それ」の指示対象は「天狗達が歌ったり踊ったりし始めた」動作を表している。このような指示現象に対しては、「始めました」という用言が表す事態が指示対象であるとして対応する。このように事態が指示対象となる場合の推定方法は、「それ」が指すものとして適正な名詞が「それ」の近くにない場合、事態が指示対象であるとして行なう(列挙判定規則 5)。

事態を指示する場合でも、次の例の「それ」のように前文の事態でなく、同一文内の逆接の接続助詞の存在する用言の表す

<sup>1</sup> これは、ここで述べている内容に対応する 6.2 節で示す規則を意味する。

事態を指示する場合がある。

おじいさんは一所懸命に歌い、そしておじいさんは一所懸命に踊りましたが、それは言葉では言い表せないほど下手糞でありました。(3)

このような場合のために、同一文内の前方に逆接の接続助詞の存在する用言があれば、その用言の表す事態を指示対象とすることにした(列挙判定規則 5)。

また、「ここ」「そこ」などの指示詞の場合、場所を指示対象に持ちやすいことから指示対象を特定することができる(判定規則 3, 4)。

太郎が公園で本を読んでいた。  
コーラを買いに売店に入りました。  
次郎はそこでで偶然、でくわしました。(4)

例えば、上の例(4)の「そこ」の指示対象は、場所を表す名詞「売店」であることがわかる。

逆に、場所を指す語が近くにない場合は接続詞であると判定する(列挙判定規則 18, 19, 20)。例えば、次の例の「そこで」は接続詞である。

歌い始めると、おじいさんは天狗が少しも怖くなくなっていました。  
そこでおじいさんは隠れていた穴から出てきてしまいました。(5)

「そこで」を英語に翻訳する際には指示詞か接続詞であるかによって“there”か“then”に訳し分けをするときにこの規則が必要である。

名詞形態指示詞は、以上で述べたもの以外に同一文内の後方照応をする場合がある。同一文内の後方照応をする場合については、他の形態指示詞も含めて文献<sup>(7)(8)</sup>での方法を利用して対処する。

### 2.2 連体詞形態指示詞の指示対象の推定方法

連体詞形態指示詞には代行指示と限定指示の二種類がある。

「この / その / あの名詞 A」<sup>2</sup>の近くに「名詞 A」があれば、限定指示と解釈して「この / その / あの名詞 A」の指示対象は「名詞 A」とする(列挙判定規則 1, 2, 3)。例えば、次の例の「このおじいさん」の指示対象はその前に出現する「おじいさん」である。

おじいさんは天狗達の前に出ていって踊り始めました。  
けれども このおじいさん は歌も踊りも下手糞でした。(6)

また、「この / その / あの名詞 A」の近くに「名詞 A」の下位語があればそれらの間に照応関係が存在すると考えられる。例えば、次の例の「あの鳥」はその前文にある「鶴」を指示している。

おじいさんは遠くの山のむこうに見えなくなってしまうまで、遠のいていく鶴の姿を見送るのでありました。  
「あの鳥を助けてやってよいことをした」とおじいさんはひとりごとを言いました。(7)

このような処理を可能にするために、本研究では「この / その / あの名詞 A」の近くに「名詞 A」の下位語がないかを調べ<sup>3</sup>、あ

<sup>2</sup> “/” は、“または”を表す。

<sup>3</sup> 語の上位下位の関係の把握は、EDR の単語辞書<sup>(9)</sup>の定義文の文末の単語をその単語の上位語とする方法<sup>(10)</sup>で行なった。

表 1: 「名詞 A の口」の用例

名詞 A になるもの
ポリ袋 ルポライター 委員長 一同 家庭教師 幹部 関係者 炎 牛 国民 採用 市民 私 自分 周作 就職 庶民 人 世間 青 木 赤ちゃん 先生 袋 谷村 担当者 炭がま 長 日本人 彼 彼 ら 彼女 被災者 避難民 負傷者 兵 兵隊 弁護士 母

ればそれを指示対象とする(列挙判定規則 1, 2, 3).

「名詞 A」やその下位語がない場合は代行指示の可能性を調べる。代行指示は指示表現が指示対象をそのまま代行するので、「その」の場合ならば「その」と置き換えて考えることができ、名詞形態指示詞による照応と同様に扱うことができる(列挙判定規則 23)。次の文の「その」の指示対象は前文の「天狗」である。

また、烏の様な顔をした天狗も居ました。  
その口はまるで烏の嘴の様に尖っているのです。(8)

また、指示詞の部分が用言の格要素になっていれば、その用言との意味的關係から指示対象をしぼることができる。だが、「その」のような連体詞形態指示詞は用言の格要素でないので用言との意味的關係から指示対象をしぼることができない。しかし、このような場合は「名詞 A の名詞 B」という用例があれば、これを用いることで指示対象をしぼることができる。たとえば、EDR の共起辞書<sup>(11)</sup>には「名詞 A の口」という用例は表 1 だけある。この用例の「名詞 A」と意味的に近いもののみが「その」の指示対象となりうるとして指示対象をしぼることができる(判定規則 5)。

また、「この」は代行指示になることはほとんどないという性質<sup>(4)</sup> <sup>(12)</sup>がある。この性質により、「この」は「名詞 A の名詞 B」の用例による検査の時、極めてよく類似した用例がない限り、代行指示ではないと判定する(判定規則 6)。

また、連体詞形態指示詞は名詞形態指示詞と同様に、前文の文末の用言が表す事態を指示する場合がある。

つまり、人間の脳より優秀なパターン認識プログラムが作れない段階では、非常に複雑で面白そうな事象については、まずその画像を作って、まずそのデータを物理学者に吟味させる必要がある。<sup>(9)</sup>  
1980 年代の初頭に LEP 実験装置の設計が始まった時、この戦略が採用されたのだった。

この例の「この戦略」の指示対象は前文の用言「吟味させる」が表す事態である<sup>4</sup>。このように事態が指示対象となる場合の推定は、代行指示として「この」が指すものや限定指示として「この戦略」が指すものとして適正な名詞が「この」の近くにない場合、事態を指示するようにすることで実現する。ただし、連体詞形態指示詞の場合も名詞形態指示詞と同様に同一文内に逆接の接続助詞の存在する用言があれば、その用言の表す事態を指示するようにしている(列挙判定規則 5)。

ところで「こんな(名詞)」については次の文を指示対象とする場合がある。

おじいさんは急に天狗達と一緒に踊りたくなってきました。とうとうおじいさんは踊りだし、踊りながら  
こんな歌を歌いました。「天狗、天狗、八天狗」<sup>(10)</sup>

<sup>4</sup> 事態を指示する場合でも代行指示か限定指示かの区別する必要があるが、本研究ではこの区別は行っていない。

表 2: 「こんな(名詞)」が前文を指示するか後文を指示するかの調査結果

「こんな名詞」につく助詞	は	はない	に	にも	には	で	では	の	すら	が	を	も	ではない
前文(個)	9	5	17	1	2	15	5	9	2	27	43	2	0
後文(個)	0	0	0	0	0	0	0	0	0	22	26	4	1

例えば、上の例の「こんな歌」の指示対象は『「天狗、天狗、八天狗」』である。ところが、「こんな(名詞)」という手がかりだけでは前方照応か後方照応かを判定することができない。そこで、「こんな(名詞)」につく助詞によって前方か後方かの判定をするために、まず、1986, 1987 年の天声人語と社説から前文か後文を指示対象とする場合の「こんな(名詞)」を抜き出して調査した。この結果を表 2 に示す。この表により、未知情報を表現する時に用いられやすい助詞「が」「を」などがつく場合以外は、前方照応であることがわかる。助詞「が」「を」がつく場合は、文献<sup>(7)</sup>の方法と同様に引用記号の“”“”がついている文が指示対象になりやすいなどの規則によって指示対象を推定する。

### 2.3 副詞形態指示詞の指示対象の推定方法

副詞形態指示詞は、名詞を指示せず文を指示する。

「天狗、天狗、八天狗」  
そう歌ったのは数えてみますとそこに八匹の天狗が居  
たからです(列挙判定規則 10, 11, 15)。<sup>(11)</sup>

例えば、この例の「そう」の指示対象は前文の『「天狗、天狗、八天狗」』である。

また、「こう」「このように」「こんなふうに」については次の文を指示対象とする後方照応をする場合がある。

天狗達は暫くおじいさんを見ていましたが、天狗達はとうとうこう言いました。  
「今日のお前は駄目だな。  
どうして昨日の様に歌ったり、昨日の様に踊ったりでき  
ないのだ。  
さあ、これを返してやるから家へ帰ってしまえ。」<sup>(12)</sup>

上の例の「こう」の指示対象は次の文以降の発話である。前方照応となる場合は文の形が「こうして」「こうすれば」などのように典型的な形になるのでこのような場合は前方照応とし(列挙判定規則 12)、これら以外の場合を後方照応とみなす。

### 3 代名詞の指示対象の推定

代名詞には人称があり、これを用いることでほぼ確実に解析することができる。会話文章の話し手や聞き手の推定は、その会話文章の発話動作を表す用言のガ格とニ格をそれぞれ話し手、聞き手とすることによって行なう。会話文章の発話動作を表す用言は、その会話文章に「と言った。」などがつけばそれとし、そうでない場合は前文の文末の用言とする<sup>5</sup>。

例えば、次の文章中の二人称の代名詞「お前さん」の指示対

<sup>5</sup> 実際にはこのような方法では発話動作を表す用言の推定を誤ることがあるが、本論文で用いたテキストではこの方法ですべて正しく解析できた。

おじいさんは地面に腰を下ろしました。
やがて(おじいさんは)眠ってしまいました。
意味素性 HUM/ANIが 眠る。
用例 彼/犬が 眠る。

図 1: 妥当性の判定の例

条件部 ⇒ { 提案 提案 .. }
提案 := ( 指示対象の候補 得点 )

図 2: 列挙判定規則の表現

条件部 ⇒ ( 得点 )
--------------

図 3: 判定規則の表現

象は、この会話文の二人称の「おじいさん」である。

「明日、また(おじいさんが)参りますよ。」とおじいさんは約束しました。  
「もちろん、お前さん(一匹の天狗が)疑う訳ではないのだが」と、一匹の天狗がおじいさんに言いました。(13)

この会話文の二人称が「おじいさん」であることは、その会話文の発話動作を表す動詞「言う」の二格が「おじいさん」であることから求まる。

#### 4 ゼロ代名詞の指示対象の推定方法

会話文章中でのゼロ代名詞の解析では、文末表現などからその動詞の省略された格要素の指示対象に入るべき人称を推定できる場合がある。このような場合は、その会話文章の一人称と二人称を推定することで代名詞と同様に会話文章中でのゼロ代名詞の解析を行なうことができる<sup>6</sup>。例えば、次の文の会話文は、「言う」のガ格と二格から一人称は「天狗達」で二人称は「おじいさん」である。

天狗達はとうとうおじいさんに言いました。  
「今日のお前は駄目だな。  
さあ、これを(天狗達が)おじいさんに返してやるから家へ(おじいさんが)帰ってしまえ。」(14)

「返してやる」のガ格の省略部分の指示対象は、「返してやる」に補助動詞「やる」がついていることから一人称の「天狗達」であることがわかる。「返してやる」の二格の省略部分の指示対象も、補助動詞「やる」から二人称の「おじいさん」と判定できる。「帰ってしまえ」のガ格の省略部分の指示対象は、命令表現から二人称の「おじいさん」と判定できる。

また、複文でのガ格の省略の場合、主節か従属節に省略されていないガ格の名詞があれば、それを他方に補完することが考えられる。ただし、文献<sup>(6)(14)(15)(16)</sup>における複文での主格の不一致の条件を満足する場合は、補完しない(列挙判定規則 8, 9)。

以上で述べた以外のゼロ代名詞については主題と焦点による選好と以下の意味素性や用例による用言との意味関係による選択制限によって求める(列挙判定規則 9, 10, 判定規則 1)。

##### ● 意味素性による判定方法

指示対象の候補となった名詞が指示対象として妥当であるための条件は、その名詞に付けられた意味素性の一つが動詞の格フレームに記述された意味素性と同一もしくは下位の意味素性であることとする。例えば、図 1 の省略部分に対しては動詞「眠る」のガ格の意味素性が HUM/ANI<sup>7</sup>であり「おじ

いさん」が HUM であることから、「おじいさん」は指示対象として妥当とする。

##### ● 用例による判定方法

指示対象の候補となった名詞と、動詞の格フレームに記述された名詞の用例とが意味的に類似していればその類似度に応じてその名詞は指示対象として妥当であるとする。例えば、図 1 の省略部分に対しては動詞「眠る」のガ格の用例が「彼/犬」であり「おじいさん」と「彼」が意味的に近いことから「おじいさん」は指示対象として妥当とする。

この用言との意味関係を用いた判定方法は、指示詞や代名詞の指示対象の推定にも用いる。

#### 5 名詞の指示対象の推定方法

「おじいさん」などの普通の名詞の指示対象の推定方法は、名詞の指示性と修飾語と所有者の三つの条件をすべて満足するときのみ照応すると判定することで実現する<sup>(1)</sup>。

#### 6 指示対象を推定する枠組

##### 6.1 推定の手順

本研究での指示詞・代名詞・ゼロ代名詞を含む名詞の指示対象の推定は、名詞の解析の手がかりとなる複数の情報をそれぞれ規則にし、これらの規則を用いて指示対象の候補に得点を与えて、合計点が最も高い候補を指示対象とすることによって実現する。

まず、解析する文章を構文解析・格解析する<sup>(17)</sup>。その結果に対して文頭から順に文節ごとにするすべての規則を適用して指示対象を推定する。規則には、指示対象の候補をあげながら候補の良さを判定する列挙判定規則とその列挙された複数の候補すべてに対して適用する判定規則の二種類がある。列挙判定規則を図 2 に、判定規則を図 3 に示す。

図中の「条件部」には文章中のあらゆる語とその分類語彙表<sup>(18)</sup>の分類番号と IPAL の格フレーム<sup>(19)</sup>の情報と名詞の指示性の情報と構文解析・格解析の結果の情報などを条件として書く。「指示対象の候補」には指示対象の候補とする単語の位置もしくは「特定指示として導入」などを書く。「特定指示として導入」のときは、個体を特定指示として新たに導入する。これは会話文中の指示詞の解析の際でテキスト中には指示対象がない場合、既出の個体を指示せず談話に新たに特定指示として個体を導入する場合に利用される。「得点」は指示対象としての適切さの度合を表す。

指示対象の推定は条件を満足した規則により与えられる得点の合計点で行なう。まずすべての列挙判定規則を適用し得点のついた指示対象の候補を列挙する。このとき同じ候補を列挙する規則が複数あれば得点を加算する。次に列挙された指示対象

<sup>6</sup> 工藤<sup>(13)</sup>は文末表現から対話コーパス中のゼロ代名詞の人称を推定しているが、本研究は物語文内の会話文章を対象としており、各発話文の話手と聞き手を推定する必要があるという点で異なる。

<sup>7</sup> HUM, ANI はそれぞれ人間(HUMAN)、動物(ANIMAL)を表す意味素性である。

表 3: 主題の重み

表層表現	例	重み
ガ格の指示詞・代名詞・ゼロ代名詞	(太郎が)した.	21
名詞は / には	太郎はした.	20

表 4: 焦点の重み

表層表現(「は」がつかないもので)	例	重み
ガ格以外の指示詞・代名詞・ゼロ代名詞	(太郎に)した.	16
名詞が / も / だ / なら / こそ	太郎がした.	15
名詞を / に / , / .	太郎にした.	14
名詞へ / で / から / より	学校へ行く.	13

の各候補に対してすべての判定規則を適用して、各候補ごとに得点を合計する。最も合計点の高い指示対象の候補を指示対象と判定する。最も合計点の高い指示対象の候補が複数ある場合は、一番初めに出された<sup>8</sup>指示対象の候補を指示対象とする。

## 6.2 指示対象の推定に用いる規則

代名詞などを解析するために列挙判定規則および判定規則をそれぞれ指示詞については 51 個と 11 個、代名詞については 4 個と 6 個、ゼロ代名詞については 24 個と 4 個作成した。主要なものを適用順序に従って以下に示す。

### 指示詞の解析のための列挙判定規則

- ソ系の連体詞形態指示詞 + 名詞 A の場合  
{ (名詞 A を部分文字列として含む名詞 45)  
(重みが  $w$  で  $n$  個前の主題で名詞 A を上位語にもつ名詞  $w - n * 2 + 10$ )  
(重みが  $w$  で  $n$  個前の焦点で名詞 A を上位語にもつ名詞  $w - n * 2 + 10$ ) }<sup>10</sup>
- コ系の連体詞形態指示詞 + 名詞 A の場合  
{ (名詞 A を部分文字列として含む名詞 45)  
(重みが  $w$  で  $n$  個前の主題で名詞 A を上位語にもつ名詞  $w - n + 30$ )  
(重みが  $w$  で  $n$  個前の焦点で名詞 A を上位語にもつ名詞  $w - n + 30$ ) }
- ア系の連体詞形態指示詞 + 名詞 A の場合  
{ (名詞 A を部分文字列として含む名詞 45)  
(重みが  $w$  で  $n$  個前の主題で名詞 A を上位語にもつ名詞  $w - n * 0.4 + 30$ )  
(重みが  $w$  で  $n$  個前の焦点で名詞 A を上位語にもつ名詞  $w - n * 0.4 + 30$ ) }
- 「それ / あれ / これ」や連体詞形態指示詞の場合で、その指示詞の直前の文節に用言の基本形か「〜とか」などの例を列挙するような表現がある場合  
{ (用言の基本形か例を列挙するような表現 40) }
- 「それ / あれ / これ」や連体詞形態指示詞の場合  
{ (前文、もしくは、指示詞の前方の同一文内に逆接接続助詞か条件形を含む用言がある場合はその用言 15) }
- 「そう / こんな / こんなに / そんなに / それほど」の解析で、これらが形容詞か副詞に係る場合  
{ (特定指示として個体を導入 30) }<sup>11</sup>
- 指示詞の場合 { (特定指示として個体を導入 10) }
- 引用文内の指示詞の場合  
{ (特定指示として個体を導入 5) }

<sup>8</sup> 規則の適用順序に従う。

<sup>9</sup> 主題が何個前かを調べる方法は、主題だけを数えることによって行なう。主題がかかる用言の位置が今解析している文節よりも前の場合はその用言の位置にその主題があると数える。そうでない場合はそのままの位置で数える。

<sup>10</sup> 列挙判定規則の提案のリストを表わす。図 2 参照。

<sup>11</sup> この規則は、柳<sup>(8)</sup>の方法を利用したものである。

表 5: 分類語彙表の分類番号の変更

意味素性	分類語彙表の分類番号	変更後の分類番号
ANI(動物)	156	511
HUM(人間)	12[0-4]	52[0-4]
ORG(組織・機関)	125,126,127,128	535,536,537,538
PLA(植物)	155	611
PAR(生物の部分)	157	621
NAT(自然物)	152	631
PRO(生産物・道具)	14[0-9]	64[0-9]
LOC(空間・方角)	117,125,126	651,652,653
PHE(現象名詞)	150,151	711,712
ACT(動作・作用)	13[3-8]	81[3-8]
MEN(精神)	130	821
CHA(性質)	11[2-58],158	83[2-58],839
REL(関係)	111	841
LIN(言語作品)	131,132	851,852
その他	110	861
TIM(時間)	116	a11
QUA(数量)	119	b11

125,126 については二つの分類番号が与えられる。

## 9. ア系の指示詞の場合

- { (特定指示として個体を導入 5) }
- 「こう / こんなふうにして / こうして」の場合 { (前文 25) }
- 「こう / こんなふうに / して / こうして」の場合 { (後文 26) }
- 「こう / こんなふうに」 + 条件節もしくは「こうして」の場合で、文末でない場合 { (前文 7) }
- 「こうだ / こんなふうだ」の場合 { (後の文 3) }
- 「ああ」の場合 { (前文 20) }
- 「そう / そうして / そのように」の場合 { (前文 30) }
- 「そう / そうして / そのように」の場合で、それらが存在する文が逆接の接続助詞か助動詞「ように」を持つ従属節である場合 { (主節 45) }<sup>12</sup>
- 「こんなに / あんなに / そんなふうに / あんなふうに」の場合で、それらが形容詞か副詞に係らない場合 { (前文 25) }
- 「ここで / そこで」の場合 { (前文 11) }
- 「ここで / そこで」の場合で、文頭の場合 { (前文 5) }
- 「ここで / そこで」の場合で、文頭であり、かつ、用言の格要素でない場合 { (前文 5) }
- 指示詞が名詞・形容詞述語文の主語の場合で、述語が「実権だ」「残念だ」「確実だ」「〜的だ」などの判断を表すものの場合 { (前文 50) }<sup>13</sup>
- 指示詞の場合で、それらが存在する文が逆接の接続助詞か助動詞「ように」を持つ従属節である場合 { (主節 10) }<sup>11,12</sup>
- 名詞形態指示詞か「その / この / あの」の場合  
{ (同一文中か前文の重みが  $w$  で主題と焦点を合わせて数えて  $n$  個前の主題  $w - n - 2$ )  
(同一文中か前文の重みが  $w$  で主題と焦点を合わせて数えて  $n$  個前の焦点  $w - n + 4$ ) }<sup>13</sup>

以上の他に表層表現から現在の時間や場所が指示対象であるとする規則などがある。

### 指示詞の解析のための判定規則

- 照応詞が名詞形態指示詞の場合で、指示対象の候補となった名詞が意味素性 HUM を満足する時、-10 点を与える。このとき指示対象の候補となった名詞の意味素性は名詞意味素性辞書<sup>(20)</sup>のものを用いる。

<sup>12</sup> この規則は、松岡<sup>(7)</sup>の方法を利用したものである。

<sup>13</sup> 指示詞の場合は、ゼロ代名詞の場合と異なり、既知情報の主題と照応するよりも未知情報の焦点と照応しやすいと考え、係数 -2, +4 により、主題よりも焦点の方が指示対象になりやすくしている。

表 6: 名詞形態指示詞の場合に与える得点

類似レベル	0	1	2	3	4	5	6	一致
得点	0	0	-10	-10	-10	-10	-10	-10

表 7: 場所を指示する指示詞の場合に与える得点

類似レベル	0	1	2	3	4	5	6	一致
得点	-10	-5	0	5	10	10	10	10

表 8: ソ系の連体詞形態指示詞の場合に与える得点

類似レベル	0	1	2	3	4	5	6	一致
得点	-10	-2	-1	0	1	2	3	4

表 9: ソ系以外の連体詞形態指示詞の場合に与える得点

類似レベル	0	1	2	3	4	5	6	一致
得点	-30	-30	-30	-30	-10	-5	-2	0

表 10: 用言との意味関係から与える得点

類似レベル	0	1	2	3	4	5	6	一致
得点	-10	-2	1	2	2.5	3	3.5	4

- 照応詞が名詞形態指示詞の場合、指示対象の候補となった名詞の分類語彙表の分類番号と以下の人間を代表する分類語彙表の番号 {5200003010 5201002060 5202001020 5202006115 5241002150 5244002100} との類似レベルの最も大きいものにより得点を与える。与える得点は表 6のとおりである。このとき用いる類似度計算には分類語彙表の分類番号を表 5に従って変換したものを用いる。この変更は、分類語彙表の分類番号の付け方が意味的に妥当でないところがあったためである。また、一桁目の数字に種類を設けたので、もとの分類番号よりも細かい分類となっている。
- 照応詞が「ここ / そこ / あそこ」の場合、指示対象の候補となった名詞が場所を意味する意味素性 LOC を満足する時、10 点を与える。
- 照応詞が「こ / そ / あそこ」の場合、指示対象の候補となった名詞の分類語彙表の分類番号と以下の場所を代表する分類語彙表の番号 {6563006010 6559005020 9113301090 9113302010 6471001030 6314020130} との類似レベルの最も大きいものにより得点を与える。与える得点は表 7のとおりである。
- 照応詞がソ系の連体詞形態指示詞の場合、それが係る名詞 B の用例「名詞 A の名詞 B」を検索し、名詞 A と指示対象の候補となった名詞の類似レベルにより得点を与える。与える得点は表 8のとおりである。
- 照応詞がソ系以外の連体詞形態指示詞の場合、それが係る名詞 B の用例「名詞 A の名詞 B」を検索し、名詞 A と指示対象の候補となった名詞の類似レベルにより得点を与える。与える得点は表 9のとおりである。ソ系以外の連体詞形態指示詞は代行指示になりにくいので、ソ系の場合よりも得点を低く設定している。
- 指示対象の候補となった名詞が格フレームの格要素の意味素性を満足しない時、-5 点を与える。
- 指示対象の候補となった名詞と格フレームの格要素の用例の名詞との類似レベルにより得点を与える。与える得点は表 10のとおりである。

#### 代名詞の解析のための列挙判定規則

- 一人称の代名詞の場合 {(一人称 25)}
- 二人称の代名詞の場合 {(二人称 25)}
- 三人称の代名詞の場合 {(一人称 -10) (二人称 -10)}
- 代名詞の場合  
 {(同一文中か前文の重みが  $w$  で主題と焦点を合わせて数えて  $n$  個前の主題  $w - n - 2$ )  
 (同一文中か前文の重みが  $w$  で主題と焦点を合わせて数えて  $n$  個前の焦点  $w - n + 4$ )}

#### 代名詞の解析のための判定規則

- 照応詞が代名詞の場合で、指示対象の候補となった名詞が意味素性 HUM を満足する時、10 点を与える。

表 11: 人称代名詞の場合に与える得点

類似レベル	0	1	2	3	4	5	6	一致
得点	0	0	3	7	10	10	10	10

- 照応詞が代名詞の場合で、指示対象の候補となった名詞の分類語彙表の分類番号と以下の人間を代表する分類語彙表の番号 {5200003010 5201002060 5202001020 5202006115 5241002150 5244002100} との類似レベルの最も大きいものにより得点を与える。得点は表 11のとおりである。

以上の他に指示詞の判定規則 7, 8 と同様な規則がある。

#### ゼロ代名詞の解析のための列挙判定規則

- 「くれる」「くださる」が補助動詞としてつく用言の格格の省略の場合で二格に省略がある場合は二格を先に解析し、格格の省略に対しては {(省略部分を埋めない -5)}<sup>14</sup>
- 引用文外で「思う」「欲しい」などの一人称が格格に入りやすい用言の格格の省略の場合、{(一人称 50)}
- 引用文中で「やる」「したい」「行く」などの一人称が格格に入りやすい用言の格格の省略の場合、{(一人称 5)}
- 「だろう」「ようだ」「そうだ」などの一人称が格格に入りやすい用言の格格の省略の場合、{(一人称 -20)}
- 引用文中で「くれる」「なせる」「来る」などの二人称が格格に入りやすい用言か命令表現か疑問表現を文末に持つ文中の格格の省略の場合、{(一人称 -30)(二人称 25)}
- 引用文中の格格の省略の場合、{(一人称 15)}
- 「A を B だと言う / 思う」における「B だ」の格格の省略の場合、{(名詞 A 50)}
- 複文での主格の不一致の条件となる接続助詞「ので」「ならば」が含まれる文で、主節(もしくは従属節)の格格が省略されていて、従属節(もしくは主節)の主格が省略されずに存在してその主格につく助詞が「が」の場合、{(従属節(もしくは主節)の主格 -30)}<sup>15</sup>
- 格格の省略の場合のデフォルト規則  
 {(重みが  $w$  で  $n$  個前の主題  $w - n * 2 + 1$ )  
 (重みが  $w$  で  $n$  個前の焦点  $w - n + 1$ )  
 (今解析している節と並列の節の主格 25)  
 (今解析している節の従属節か主節の主格 23)  
 (今解析している節が埋め込み文の場合で主節の主格 22)}
- 格格以外の省略の場合のデフォルト規則  
 {(重みが  $w$  で  $n$  個前の主題  $w - n * 2 - 3$ )  
 (重みが  $w$  で  $n$  個前の焦点  $w - n * 2 + 1$ )}
- 今求める省略要素を格要素に持つ用言の他の格要素に名詞 A がすでに入っている場合、{(名詞 A -20)}

以上の他に名詞述語文の述語に来る名詞から前文が指示対象である推定する規則がある。

#### ゼロ代名詞の解析のための判定規則

- ゼロ代名詞の解析にも指示詞の判定規則 7, 8 と同様な規則がある。

#### 名詞の解析のための規則

名詞の解析は列挙判定規則のみで行なった。列挙判定規則は 9 個作成したが、そのうちの主要なものを適用順序に従って以下に示す。詳細は文献<sup>(1)</sup>を参照。

#### 名詞の解析のための列挙判定規則

- 推定した名詞の指示性が定名詞の場合で、その名詞を末尾に含み修飾語や所有者が同じ名詞 A が前方にある場合  
 {(名詞 A 20)}
- 名詞の指示性が総称名詞の場合  
 {(総称指示として個体導入 10)}
- 名詞の指示性が総称名詞でも不特定性の不定名詞でもない場合  
 {(特定指示として個体導入 10)}

#### 6.3 指示対象の推定例

指示詞「この」の指示対象を推定した例を図 4 に示す。これは図中の下線部の「このドル高」の指示対象を前文と正しく解

<sup>14</sup> この規則は視点の理論<sup>(3)</sup>を利用したものである。

<sup>15</sup> この規則は平井<sup>(16)</sup>の方法を利用したものである。

表 12: 本研究の実験結果

テキスト	文数	名詞	指示詞	代名詞	ゼロ代名詞
学習サンプル	204	85%(130/153)	87%( 41/ 47)	100%( 9/ 9)	86%(177/205)
テストサンプル	184	77%( 89/115)	86%( 42/ 49)	82%( 9/11)	76%(159/208)

各規則で与える得点は学習サンプルにおいて人手で調節した。学習サンプルは、例文(43文)、童話「こぶとりじいさん」全文(93文)<sup>(21)</sup>、天声人語一日分(26文)、社説半日分(26文)、サイエンス(16文)であり、テストサンプルは、童話「つるのおんがえし」前から91文抜粋<sup>(21)</sup>、天声人語二日分(50文)、社説半日分(30文)、サイエンス(13文)である。

ドル相場は、米新政権の経済政策に対する期待の高まりなどから130円台に上昇した。 このドル高は、米国と欧州各国との間の政策協調をぎくしゃくさせている。					
規則	各候補の得点(点)				
	前文	特定指示として導入	130円台	高まり	ドル相場
列5	15				
列7		10			
列23			17	15	13
判6			-30	-30	-30
合計	15	10	-13	-15	-17
“列5”などは指示詞の列挙判定規則5などを意味する。					
図 4: 指示詞「この」の指示対象の推定例					

析したことを示している。

## 7 実験と考察

指示対象の推定を行なう前に構文解析・格解析を行なうが、その際の誤りは人手で修正した。格フレームはIPALの辞書のものを用いたが、IPALの辞書にない用言に対しては人手で格フレームを作成した。本研究による方法で名詞、指示詞、代名詞、ゼロ代名詞の指示対象を解析した実験結果を表12に示す。指示詞の解析において、指示対象が文になる場合で、前の数文もしくは後ろの数文が指示対象となる場合は、その範囲まで推定できなくても、指示対象が前方の文か後方の文かの判定を行なうことができれば正解とした。これは指示対象の範囲の推定は、原因-結果、例提示などの文の間の関係を解析した後ですべきであると考えためである。また、ゼロ代名詞の解析精度は指示対象が存在するか否かがあらかじめわかっていると仮定して解析した時の精度である。

### 7.1 指示詞の実験に対する考察

正解率はテストサンプルにおいても80%を越えたので、本システムで用いた規則は有効であることがわかる。しかし、指示詞は種類が多いのでより詳細に規則を作成することでさらに高精度に解析できる可能性がある。また、本研究では「この」が代行指示になりにくいという性質を利用したが、これを利用したために正しく解析できた例が4例あった。

指示詞の解析を誤ったものとして次のようなものがあった。

けれども、もしおじいさんが町へたきぎを売りに行かなければ、自分と妻が食べるためのお米を買うお金がないです。  
そこでおじいさんは寒さも雪もなんのその、山道をとぼとぼ歩いていきました。(15)

この例の「そこで」は接続詞として用いられているが、前文

に場所を意味する「町」がありこれを指示対象と誤って解析した。

### 7.2 代名詞の実験に対する考察

今回の実験テキストでは、一人称と二人称の代名詞しか出現しなかったため、その代名詞が存在する会話文の一人称と二人称を推定することでほぼ正確に推定できた。

代名詞の解析において誤った例は、すべて次のように会話文の二人称の推定を誤ったためであった。

老夫婦は娘がそう話している間も寒さで震えているのが目にうつりました。  
娘の頼みごとを聞いておじいさんは(娘に)言いました。  
「それはなんとも気の毒だ。  
もちろん、おまえさんを泊めてあげるのはかまわないよ。  
なあ、ばあさんや」(16)

例えば、この例の二人称の代名詞「おまえさん」の指示対象は「娘」である。ところが、「おじいさんは(娘に)言いました。」の「言う」の二格の省略部分の解析を「老夫婦」と誤ったためにこの会話文の二人称を「老夫婦」と誤って解析し、「おまえさん」の指示対象を「老夫婦」と誤って解析した。この誤りは「おまえさん」が単数であることから「言う」の二格に補充されるものが単数であるものに限ることに対処できる。

### 7.3 ゼロ代名詞の実験に対する考察

ゼロ代名詞の解析誤りとしては、分類語彙表、意味素性辞書、格フレーム辞書に誤りがあるために誤ったものや、統語構造や補助表現から指示対象が推測できるのに規則が不十分のために誤ったものがあった。

また、理解や推論が必要なために誤ったものとして次のものがある。

そんな状況なのに、ワシントンで開かれる主要先進7カ国の蔵相中央銀行総裁会議(G7)について各国の通貨当局は「大きな問題はないので共同コミュニケは出ない。顔合わせ中心の会合だ」と、まるで会議の意義を薄めようとしているような言い方だ。  
(中略)

米新政権は近く、財政赤字削減の具体的構想を議会に示す予定である。(17)

(段落替え)  
(通貨当局が)共同コミュニケの発表を控えるのは、為替市場に過大な期待を与えたくないためだろう。

この例の「控える」のガ格の省略部分の指示対象は「各国の通貨当局」である。しかし、システムは「米新政権」を誤って指示対象と解析した。この指示対象を正しく解析することができるようにするためには、そこまでの文章から共同コミュニケの発表を控えるものが通貨当局であることを理解しておく必要がある。

表 13: 意味素性と用例の対照実験の結果

照応詞	テキスト	用例と意味素性の両方を用いる	意味素性のみを用いる	用例のみを用いる (分類語彙表の分類番号の変更)	用例のみを用いる (分類語彙表の分類番号のまま)	意味素性と用例の両方を用いない
指示詞	学習サンプル	87% (41/47)	83% (39/47)	87% (41/47)	83% (39/47)	79% (37/47)
	テストサンプル	86% (42/49)	88% (43/49)	88% (43/49)	84% (41/49)	86% (42/49)
代名詞	学習サンプル	100% (9/9)	100% (9/9)	100% (9/9)	100% (9/9)	89% (8/9)
	テストサンプル	82% (9/11)	64% (7/11)	82% (9/11)	55% (6/11)	64% (7/11)
ゼロ代名詞	学習サンプル	86%(177/205)	83%(171/205)	86%(176/205)	82%(169/205)	66%(135/205)
	テストサンプル	76%(159/208)	76%(158/208)	79%(164/208)	75%(155/208)	63%(131/208)

#### 7.4 対照実験

指示詞と代名詞とゼロ代名詞の解析において判定規則として用例を用いる規則と意味素性を用いる規則とを用いるが、これらの有効性を調べるために表 13にあげる対照実験を行なった<sup>16</sup>。表のように用例を用いる方法と意味素性を用いる方法とは同程度の精度であった。このことにより、意味素性と同様に用例を用いることができることがわかった。また、分類語彙表においては分類番号の付け方が意味的に妥当でないところがあったため分類番号を変更したが、分類番号を変更したものが分類語彙表の分類番号のままのものより精度が良く、妥当な変更であることが確認できた。

実験結果を考察すると用例による方法では格フレームの記述から外れた表現に対しても有効な場合があった。例えば、「言う」の二格には人間しか入らないように格フレームに書いてあるので、意味素性による方法では次の例の二格の省略部分には鶴を補うことができない。

おじいさんは鶴をはなしてやりながら (鶴に) 言いました。 (18)

しかし、用例による方法では人間と動物は表5により類似レベルが1で減点が小さく鶴を二格に補うことができる。

#### 8 おわりに

対照実験を通じて意味的制約として意味素性と同様に用例を用いることができることがわかった。また、連体詞形態指示詞の指示対象の推定に意味的制約として「名詞 A の名詞 B」の用例を利用するなどの新しい手法を提案した。

#### 参考文献

(1) 村田真樹, 長尾真, 日本語文章における名詞の指示対象の推定, 言語処理学会第1回年次大会発表論文集, A4-4, (1995).

(2) 長尾真, 辻井潤一, 田中一敏, 意味および文脈情報を用いた日本語文の解析 — 文脈を考慮した処理, 情報処理学会誌, Vol. 17, No. 1, (1976), pp. 19-28.

(3) Megumi Kameyama, A Property-sharing Constraint in Centering, 24th Annual Meeting of the Association for Computational Linguistics, (1986), pp. 200-206.

<sup>16</sup> ここでいう用例を用いる規則は指示詞の判定規則2, 4, 代名詞の判定規則2および指示詞の判定規則8とそれに対応する代名詞・ゼロ代名詞の判定規則を意味する。また、意味素性を用いる規則は上の用例を用いる規則に対応する意味素性を用いる規則を意味する。「名詞 A の名詞 B」の用例を用いる規則には対応する意味素性を用いる規則がないので、この対照実験ではすべての場合でこの規則は用いた。

(4) 山村毅, 大西昇, 杉江昇, 日本語指示詞の前方照応現象の分類, 信学論, Vol. J75-D-II, No. 2, (1992), pp. 371-378.

(5) Shingo Takada, Norihisa Doi, Centering in Japanese: A Step Towards Better Interpretation of Pronouns and Zero-Pronouns, COLING '94, (1994), pp. 1151-1156.

(6) 中岩浩巳, 池原悟, 語用論的・意味論的制約を用いた日本語ゼロ代名詞の文内照応解析, 言語理解とコミュニケーション研究会 NLC95-5, (1995), pp. 33-40.

(7) 松岡正男, 村田真樹, 黒橋禎夫, 長尾真, 表層表現を利用した日本語文章における後方照応表現の自動抽出, 情報処理学会自然言語処理研究会 95-NL-108, (1995).

(8) 柳喜芳, 主題・焦点と表層表現に着目した照応処理, 京都大学工学部修士論文, (1994).

(9) (株)日本電子化辞書研究所, EDR 電子化辞書 日本語単語辞書評価版第2.1版, (1994).

(10) 鶴丸弘昭, 竹下克典, 伊丹克企, 柳川俊英, 吉田将, 国語辞典を用いたソーラスの作成について, 情報処理学会研究報告, Vol. 91, No. 37 (91-NL-83), (1991).

(11) (株)日本電子化辞書研究所, EDR 電子化辞書 日本語共起辞書評価版第2.1版, (1994).

(12) 正保勇, 「コンア」の体系, (国立国語研究所, 1981).

(13) 工藤育男, 友清睦子, 日本語の述部の特性を用いた省略の補完機構について, 信学論, Vol. J76-D-II, No. 3, (1993), pp. 624-635.

(14) 南不二男, 現代日本語の構造, (大修館書店, 1974).

(15) 吉本啓, 談話処理における日本語ゼロ代名詞の扱いについて, 自然言語処理研究会 86-NL-56, (1986).

(16) 平井誠, 省略語と事象間関係から見た日本語文の意味解析と文脈解析の枠組みについて, 言語理解とコミュニケーション研究会 NLC86-8, (1986), pp. 7-14.

(17) S. Kurohashi and M. Nagao, A method of case structure analysis for Japanese sentences based on examples in case frame dictionary, Vol. E77-D, No. 2, (1994), pp. 227-239.

(18) 国立国語研究所, 分類語彙表, (秀英出版, 1964).

(19) 情報処理振興事業協会技術センター, 計算機用日本語基本動詞辞書 IPAL(Basic Verbs) 説明書, (1987).

(20) 渡辺靖彦, 黒橋禎夫, 長尾真, IPAL 辞書と分類語彙表を用いた単語意味辞書の作成, 情報処理学会第45回全国大会予稿集, 6F-8, (1992).

(21) 中尾清秋, こぶとりじいさん 他 鶴の恩がえし, きき耳ずきん, 英訳「日本むかしばなし」シリーズ, 第7巻, (日本英語教育協会, 1985).