

日本語文章の複文におけるゼロ代名詞照応の解析に関する一考察

桃 内 佳 雄*

A Study on the Resolution of Zero Pronouns in Japanese Complex Sentences

Yoshio MOMOUCHI*

あらまし

文章理解過程モデルの構成の中で、文章中に現れる代名詞による照応の解析アルゴリズムは一つの重要な構成要素である。本論文では、日本語文章の複文におけるゼロ代名詞による照応の種類とその解析のための基本的なアルゴリズムについて考察する。ゼロ代名詞は、もし表現するとすれば、名詞に対応する表現の省略であり、その解析は、文脈や状況の中に存在する情報を基本的な手がかりとして、語レベルの情報から、文の統語構造、意味構造、さらには語用論的な情報も利用して行われなければならない。特に、複文におけるゼロ代名詞照応の解析アルゴリズムは、複文の特徴や構造を十分考慮に入れて構成されなければならない。本論文では、まず、小学国語教科書文章中の複文に現れる具体的なゼロ代名詞照応の分析に基づいて、複文におけるゼロ代名詞照応の基本的な類型をまとめる。次に、照応解析アルゴリズムの一つとして提案されているセンター理論に基づくアルゴリズムの拡張について検討し、複文の特徴や構造を考慮にいたしたゼロ代名詞照応解析のための基本的なアルゴリズムの構成と具体例への適用について考察する。

1. はじめに

文章の理解において、文章の正しい意味を理解するためには、文脈情報と様々な知識の利用

* 北海学園大学工学部電子情報工学科

* Department of Electronics and Information Engineering, Faculty of Engineering, Hokkai-Gakuen University

が必須である。特に、代名詞による照応関係や文あるいは節間の接続関係の解析においては、文脈情報の利用のしくみを明らかにすることが重要な課題である。本論文では、日本語文章の複文におけるゼロ代名詞照応の類型とその解析のための基本的なアルゴリズムについて考察を進める。ゼロ代名詞は、もし表現するとすれば、名詞に対応する表現の省略であり、その解析は、文脈や状況の中に存在する情報を基本的な手がかりとして、語レベルの情報から、文の統語構造、意味構造、さらには、語用論的な情報も利用して行われなければならない。特に、複文におけるゼロ代名詞による照応の解析は、複文の構造を十分考慮に入れて行う必要がある。

すでに、日本語文章における省略現象一般については、日本語学において多くの研究[7,11,13,17]が行われており、森岡[17]は省略を次のように規定して考察を進めている。(1) 言語の省略とは、話し手、聞き手あるいは観察者によって、ある要素が省かれていると実感される現象である；(2) 省略されていると意識される要素は、言語的あるいは非言語的文脈から復元可能でなければならない；(3) 省略されていると意識される要素は、他の言語表現と比較することによって生ずる。ここでは、省略の同定(実感)、省略要素の復元、文脈情報の参照が基本的な要素過程として設定されている。

ゼロ代名詞照応に関する研究は、言語学的な考察に加えて、認知科学的あるいは計算言語学的な視点からの多くの考察が行われており[1,16 およびその参考文献参照]、焦点やセンターなどの文脈情報の選択的、構造的な保持機構の利用を基本として、文の表層的な情報や意味情報、さらには語用論的情報を制約として利用する解析手法が提案されている。それらの研究の中で、複文におけるゼロ代名詞照応とその解析アルゴリズムに関する研究も進められている[9,18,19,20,24,26]。複文は、複数の節から構成され、文章の中で、構造を持ったひとつのまとまりとして、他の単文や複文と様々な関係で関連しつつ存在している。従って、複文におけるゼロ代名詞照応解析アルゴリズムは、そのような複文内部の構造や前後の文との関連構造への考慮を組み込んだ形で構成されなければならない。

本論文では、まず、小学国語教科書文章中の複文中に出現するゼロ代名詞による照応現象を分析し、複文におけるゼロ代名詞照応の基本的な類型についてまとめる。次に、センター理論に基づく照応解析アルゴリズムの拡張について検討し、複文の特徴や構造を考慮にいたしたゼロ代名詞照応解析のための基本的なアルゴリズムの構成とその具体例への適用について考察する。実例の出典資料は、「教育出版株式会社発行：しょうがくこくご／小学国語」である。

2. 複文におけるゼロ代名詞による照応

2.1 ゼロ代名詞による照応の基礎

ゼロ代名詞は、もし表現するとすれば名詞に対応する表現の省略形である。例えば、次のよ

うな複文の例について考えてみよう。

<1>みさがそばへ行って、子犬をだき上げると、小さなしっぽをこまかくふって、みさのかおをなめました。

この文は、4つの節から構成される複文である。最初の節を除く3つの節で、主格（が格）に対応する表現が省略されている。省略されていることを次のように表現する。

<1'>みさがそばへ行って、[φ1が]子犬をだき上げると、[φ2が]小さなしっぽをこまかくふって、[φ3が]みさのかおをなめました。

φ1, φ2, φ3がゼロ代名詞の表現である。このゼロ代名詞に対応する名詞表現を復元すると次のようになるであろう。

<1''>みさがそばへ行って、[みさが]子犬をだき上げると、[子犬が]小さなしっぽをこまかくふって、[子犬が]みさのかおをなめました。

ここで、例えば、2番目の節、「[みさが]子犬をだきあげると、」において、「みさが」が省略されているという実感は、表層的には、「だき上げる」という動詞の表層格である「主格(が格)」に対応する表現が欠けているという認識、意味的には、「だき上げる」という行為の必須的な深層格である「動作主」に対応する要素が欠けているという認識から生ずると考えられる。そして、それは一つ前の1番目の節の主格表現「みさが」を参照して復元される。また、上の例では、[φ2が]、[φ3が]に対応する復元形は、[子犬が]である。それぞれ、「ふる」、「なめる」という動詞が意味する行為の必須的な深層格に対応する。これは、2番目の節中の「子犬を」という表現を参照して復元される。この場合、「子犬」が2番目の節で担う役割と、3番目、4番目の節で担う役割が異なっている。

以上より、ゼロ代名詞の復元に至る基本的な過程を次のようにまとめることができる。

(1) ゼロ代名詞の同定

動詞が意味する行為にとっての必須格要素が欠けているということを同定する。

<1> で、動詞「だき上げる」の必須格フレームは次のようであるとする。

(だき上げる	:	(動作主	<人間>	が)	(対象	<物>	を))
		深層格	意味特徴	表層格	深層格	意味特徴	表層格	

2番目の節では、「だき上げる」の必須格要素（動作主 <人間> が）が欠けている。このような格フレーム情報を知識として持ち、利用することによって同定が可能になる。

(2) ゼロ代名詞の照応要素の探索

ゼロ代名詞が参照する要素（照応要素）を、ゼロ代名詞を含む節からの制約を考慮しながら、文脈または外界の情報の中に探索する。

<1>の最初の節から導入される文脈情報は例えば次のようである。これは、最初の節で名詞により導入された指示対象のリストである。意味特徴と格情報も付与してある。

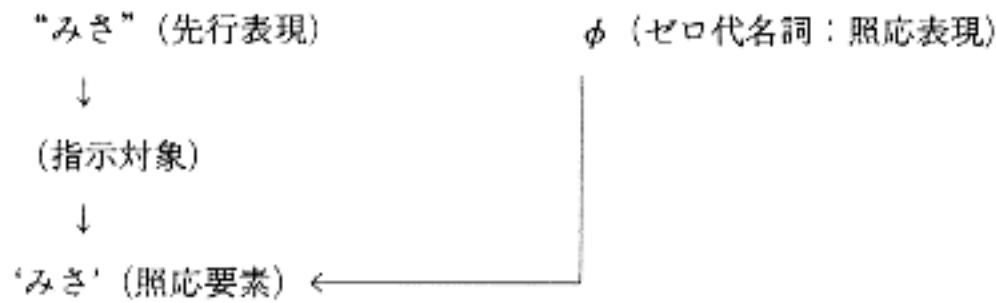
((‘みさ’, “みさ”, 人間, 動作主/が) (‘そば’, “そば”, 位置, 場所/に))

“みさ”は表現, ‘みさ’は指示対象を表わすものとする。制約を考慮しながら上のリストを探索し, ゼロ代名詞 $\phi 1$ の照応要素を ‘みさ’, 復元形を “みさが” とする。制約は, 例えば, 次のように表現することができる。

[($\phi 1$ <人間> <が:動作主>) [子犬をだきあげると]]

これは, ゼロ代名詞を含む節からの可能な制約のすべてを意味していると考え。基本的には, [<子犬をだきあげることのできる<人間>> で, <が格:動作主> となりうる要素] として, 文脈リストの探索を制約することになる。

この例では, <1>において, 照応要素を最初の節の “みさ” という表現の指示対象 ‘みさ’ として同定している。文脈中に照応要素となる対象を導入するもとなった表現を先行表現と呼ぶことにする。ただし, そのような表現は後続する文脈中に存在する場合もあるが, そのような場合も含めて先行表現と呼ぶことにする。また, 一般に, ゼロ代名詞も含めて, 代名詞などの照応する側の表現を照応表現と呼ぶ。例<1>での照応表現(ゼロ代名詞), 先行表現, 照応要素の関係は次のようになる。本論文では, ゼロ代名詞による照応の解析をゼロ代名詞が参照する照応要素を探索することと考える。



引き続き, 同じように $\phi 2$, $\phi 3$ の解析を進めて, <1>を得る素朴な手順を考えてみよう。文脈リストは, 前の節からの情報のみを保持し, 要素の並びは, 節における左から右への出現順であり, その中の探索は, 先頭から行うことにする。制約は, ゼロ代名詞を含む節からの可能な制約を考慮する。このような素朴な手順で, 例<1>については, すべてのゼロ代名詞の解析に成功する。2番目, 3番目の節の解析後の文脈リストとゼロ代名詞探索における制約は次のようになるであろう。 $\phi 1$, $\phi 2$ が復元されたところで, $\phi 1$, $\phi 2$ を含む節からの文脈リストがそれぞれ構成される。また, 例えば, 2番目の節からの制約は, [<小さなしっぽをこまかくふる<動物>> で, <が格:動作主> となりうる要素] として, 文脈リストの探索を制約する。

((‘みさ’, “みさ”, 人間, 動作主/が) (‘子犬’, “子犬”, 動物, 対象/に))

[($\phi 2$ <動物> <が:動作主>) [小さなしっぽをこまかくふって]]

((‘子犬’, “子犬”, 動物, 動作主/が) (‘しっぽ’, “しっぽ”, 動物の部分, 対象/を))

[($\phi 3$ <動物> <が:動作主>) [みさのかおをなめました]]

2.2 ゼロ代名詞による照応の基本的な分類

本論文で考察する複文は、益岡・田窪 [12] による次のような構成と分類に基づいて考える。

[複文は複数の節で構成されるわけであるが、それらの中で原則として、文末の述語を中心とした節が文全体をまとめる働きをする。この節を「主節」と呼ぶ。主節以外の節は、主節に対して特定の関係で結びつく。これらの節を一括して「接続節」と呼ぶ。・・・接続節は、主節に対する関係の違いによって、「従属節」と「並列節」に分けられる。[12: pp. 4 - 5]]

さらに、従属節は、補足節(引用節も含む)、副詞節、連体節に分けられる。本論文では、接続節を副節と呼ぶことにする。前節の例<1>における節の構成は次のようになる。

<1>みさがそばへ行って、	[副節：並列節]
子犬をだき上げると、	[副節：副詞節]
小さなしっぽをこまかくふって、	[副節：並列節]
みさのかおをなめました。	[主節]

複文におけるゼロ代名詞照応の基本的な分類は、まず、一般的なゼロ代名詞照応の基本的な分類 [16, 19] に対応して、照応要素を同定するための基本的な手がかりである先行表現が文章の中に存在する場合(文章内照応)と存在しない場合(文章外照応)とに分け、さらに、文章内照応について、先行表現がゼロ代名詞を含む複文内にある場合(文内照応)と複文外の文にある場合(文間照応)に分けるという分類である。文章外照応には、書き手(作者)への照応、一般的(総称的)な対象への照応などがある。さらに、照応要素が、先行表現から直接導入される要素か、推論により間接的に導入される要素かという分類の視点も考えられる [29]。前節の例<1>中のゼロ代名詞は、すべて文内照応の例である。上の基本的な分類に基づいて、小学国語テキストベース1~3年(ET 13 コーパス)からゼロ代名詞の出現例を示す。

<1>文内照応

<2>おじいさんは、[φ1が]うれしくなって、[φ2が]おむすびをみんなあなにおとしました。

・文頭に主題表現が一つあり、複文中のすべての節の主格がゼロ代名詞化されて、その照応要素は、主題表現の指示対象である主題要素である。

<3>ろくべえは、[φ1が]シャボン玉を見ると、[φ2が][φ3を]食べものとまちがえるのか、[φ4が][φ5に]すぐとびつきます。

・φ3、φ5の照応要素は、「シャボン玉」である。各節の中での役割は必ずしもすべて同じではない。主格以外のゼロ代名詞の例である。

<4>わたしは、[φ1が]ミーコの大活やくの話を知ったので、[φ2が]画用紙で「ひょうしゅうじょう」を作って、[φ3が][φ4に][φ5を]魚といっしょにやりました。

・φ4の照応要素「ミーコ」は、格要素に対応する名詞句「ミーコの大活やくの話」中の「ミーコ」

の指示対象である。また、それは直前の節からの文脈リストにも含まれていない。

<5>「ねえ、この森ん中をたんけんしようよ。」[φ1が]こんなことを言って、この森の中に、
達夫とちよが入っていったのは、三日ばかり前のことでした。

・φ1の先行表現“達夫とちよ”は、後続文脈中の節の主格である。さらに、前の二つの節は二つあわせて一つの補足節を構成している。

<II>文間照応

<6>あげはちょうが、みかんのきにとんできました。

[φ1が]ときどきおなかのさきをまげて、[φ2が]はに、なにかつけています。

・主格ゼロ代名詞φ1、φ2ともに、先行表現は前文の“あげはちょう”である。文間照応における先行表現は、必ずしも前文中にあるわけではない。藤澤・増山・内藤 [4] らによる分析では、約87%が2文前までにあるということが示されている。

<7>モユちゃんは、[φ1が]友子にピシャピシャ顔をたたかれながら、[φ2が]きょとんと友子を見上げていました。[φ3が]それを見ると、友子は、[φ4が]なみだをぼろぼろこぼしながら、[φ5が]モユちゃんにほおずりました。

・φ1、φ2ともに、先行表現は、文中の主題表現“モユちゃん”である。第2文では、主題表現が、文頭ではない位置に現れている。φ3をどのように解析するかが問題である。

<8>ある日、トンキーとワンリーが、ひょろひょろと体をおこして、[φが]ぞうがかりの前にすすみ出てきました。

[φが]おたがいに、ぐったりとした体をせなかでもたれあって、[φが]げいとうをはじめたのです。

[φが]後ろ足で立ち上がりました。

[φが]前足を上げておりまげました。

[φが]はなを高く上げて、[φが]ばんざいをしました。

[φが]しなびきった体じゅうの力をふりしぼって、[φが]よろけながら[φが]いっしょうけんめいです。

[φが]げいとうをすれば、[φが]もとのようにえさがもらえとおもったのでしょうか。

・第2文以下には、単文と複文が混在しているが、第1文の表現“トンキーとワンリー”を先行表現として、その指示対象がすべてのゼロ代名詞(φ)の照応要素となっている。すべて、主格ゼロ代名詞で、物語文章では、物語の登場人物を主格とする文の連鎖(継続)がしばしば起こり、対応して、この例のような主格ゼロ代名詞の連鎖が起こる。

<III>文章外照応

<9>日よう日のあさ、べんきょうがおわったので、[φが]そとへ出てみました。

・作文文章の最初の文である。ゼロ代名詞の照応要素は、書き手である‘わたし’である。

<10>「ふかいあなのそこには、ガスがたまっていて、[φ1が]それをすうと、[φ2が]しぬことだってあるんですよ。」

・この例のゼロ代名詞の照応要素は、一般的、総称的な[だれか(人)]である。あるいは、[人]に限らず、[生き物一般]といったほうがよいかもしれない。

<11>こんど、みさがひっこしてきたのは、大きなだんちの三がいで、[φが]犬をかうことはできません。

・この例のゼロ代名詞の照応要素も、一般的、総称的な[だれか(人)]である。上の2つの例は、文そのものが一般的な記述であるということを同定できるということを前提として、解析されることになるであろう。その手がかりとして、上の二つの例について、「～すること」という補足節の主格であるということが考えられる。

2. 3 副節におけるゼロ代名詞による照応

前節では、ゼロ代名詞に対する先行表現が、そのゼロ代名詞が出現する複文の内にあるか外にあるかという視点から、複文におけるゼロ代名詞照応について検討した。本節では、ゼロ代名詞が出現する副節の種類に着目して、複文におけるゼロ代名詞照応について検討する。副節の種類は、補足節(引用節も含む)、連体節、副詞節、並列節である。副節の種類によって、ゼロ代名詞の出現、ゼロ代名詞の照応表現の存在位置などに相違が見られるであろうか。

(1) 補足節(引用節も含む)

<12>ぼくが学校から帰ると、父は、[[φが]長いすでねむっていること]があります。

・照応表現が複文内の主題表現で、補足節の直前にある。

<13>こんど、みさがひっこしてきたのは、大きなだんちの三がいで、[[φが]犬をかうこと]はできません。

・これは、前節でも例としたが、文章外の照応要素は一般的(総称的)な対象である。

補足節に現れるゼロ代名詞は、主格ゼロ代名詞がほとんどで、主題表現が存在すれば、その主題要素を照応要素とすることが多く、また、文内照応が多い。しかし、<13>のような文章外照応の例もある。

(2) 連体節

<14>ぼくは、[[φ1が][φ2を]学校に通う道]のことを思い出しました。

<15>夏のはじめ、町に植木市がたった時、女の子は、[[φ1が][φ2を]ためておいたおこづかい]を全部使って、なえ木を二本買いました。

・上の2つの例では、主格ゼロ代名詞φ1に対する照応要素は複文内の主題要素である。また、被修飾名詞に対応する要素が「を格ゼロ代名詞」φ2の照応要素となっている。連体節に現われる「を格ゼロ代名詞」の照応要素として、被修飾名詞に対応する要素が有力な候補となる。主格

ゼロ代名詞については、必ずしも、そうではない。

(3) 副詞節

<16> モユちゃんは、[φ1が] 友子にピシャピシャ顔をたたかれながら、[φ2が] きょとんと友子を見上げていました。

<17> どちらも、[φ1が] てつのおりにもたれ、[φ2が] はなを長くのばして、[φ3が] ばんざいのげいとうをしたまま、[φ4が] しんでしまいました。

<18> さや子は、「こわい、こわい。」とさわぎました。

[φ1が] あんまりさわぐので、みのわくんがかい中電とうをつけました。

<19> [φ1が] うちにはいってから、おじさんは、[φ2が] しずかにかさをつぼめました。

<20> ねずみは、[φ1が] あと足だけで立とうとするとき、[φ2が] しっぽをまっすぐのばして、[φ3が] からだのつりあいをとります。

・主題要素、あるいは主格要素への照応が多い。接続語句によって、ゼロ代名詞の現われ方に相違が見られる。そのことについては、まとめて後述する。

(4) 並列節

<21> 友子は、[φ1が] 夜中に何度も起きて、[φ2が] 物おきのすみにあるモユちゃんのわらのベッドをのぞきに行きました。

<22> 止まったちょうは、[φ1が] 長い口をのばして、[φ2が] みつをすおうとしています。

<23> 先生は [φ1が] さっそく二人をよんで、[φ2が] ききました。

・すべての主格ゼロ代名詞について、照応要素が複文内の主題要素である。

(5) 副詞節と並列節

<24> わたしは、[φ1が] ミーコの大活やくの話を聞いたので、[φ2が] 画用紙で「ひょうしょうじょう」を作って、[φ3が] [φ4に] [φ5を] 魚といっしょにやりました。

・主格ゼロ代名詞について、照応要素が複文内の主題に対応する要素である。

<25> [φ1が] 外へ出ようと思って、[φ2が] げんかんの戸を開けると、かわいい子ねこがちよこんとすわっていました。わたしは、[φ3が] びっくりしましたが、[φ4が] 寒そうにしているので、[φ5が] [φ6を] 家の中に入れてやりました。

・第2文の「ので節」の主格ゼロ代名詞 φ4 は、複文外の前の文の主節の主格要素「かわいい子ねこ」を照応している。

野田 [21] は、従属節（本論文における従属節と並列節）と主節とからなる複文の「は」と「が」について、次のような特徴をまとめている。

	従属節と主節の主格が違う	従属節と主節の主格が同じ
従属句	① ×	⑤ ~は/が [...ながら] …。
強い従属節	② [~が…たら], ~は/が…。	⑥ ~は [...たら], …。

弱い従属節 ③ [～は／が…けれど], ～は／が…。 ⑦ [～は…けれど], …。

引用節 ④～は／が, [～は／が…と] …。 ⑧ ～は, […と] …。

従属句 : 付帯状況句 (～ながら, ～まま, ～て)

継起句 (～て, ～[連用形])

○「～は」と「～が」を含むことができない。

強い従属節: 継起節 (～と, ～たら, ～て, ～[連用形])

仮定節 (～たら, ～(れ)ば, ～と, ～ては, ～ても)

様態節 (～ように, ～ほど)

時間節 (～とき, ～まえに, ～あとで, ～まで)

連体修飾節 (～[名詞])

名詞節 (～こと, ～の, ～か)

理由節 (1) (～ため, ～て, ～から (焦点), ～ので (焦点), ～のに (焦点))

○「～は」を含むことができない。

弱い従属節: 理由節 (2) (～から, ～ので, ～のに)

並列節 (～て, ～[連用節], ～し, ～けれど, ～が)

引用節 : 引用節 (～と, ～って)

この分類は、南 [14,15] による従属句の分類をもとにしており、上の表の「従属句」が南による「A類の従属句」に、「強い従属節」が「B類」に、「弱い従属節」が「C類」にほぼ対応している [21]。また、主題表現「～は」と主格表現「～が」の特徴付けについても、ほぼ一致している。以下では、上の野田による特徴づけに基づいて考えることにする。次の従属句 (～ながら) を含む例について、考えてみよう。

<26> モユちゃんは、[φ1が] 友子にピシャピシャ顔をたたかれながら、[φ2が] きょとんと友子を見上げていました。

野田による上のまとめによれば、従属句は、「～は」と「～が」を含むことができない。このことは、従属句の主格は、つねにゼロ代名詞化されて、従属句の外にその照応要素を探さなければならないということの意味し、しかも、その照応要素は、従属句にとって最も近くにある主格となりうる要素であるとするのが妥当な考えであるということになる。従って、上の例では、ゼロ代名詞 φ1 の照応要素は、「モユちゃん」であるとするのが自然な解釈である。ここでの議論は、文献 [16] で考察された接続語句制約 [複文の文頭で導入される主題は、その複文中の従属句 (南の A 類) および強い従属節 (南の B 類) の主格ゼロ代名詞の照応要素とはならない。] という制約の誤りを改めるものとなっている。すなわち、次のように改める。

【接続語句による制約規則: 従属句の主格ゼロ代名詞の照応要素は、その従属句を含む複文中

の最も近い主格要素あるいは主題要素である。】

このような接続語句による制約規則については、吉本 [30]、田村・奥村 [26]、中岩・池原 [19] による考察と実験がある。田村・奥村 [26] によってまとめられている規則を以下に示す。ここで、前、後というのは接続語句の前の節、後の節をさしている。主語の一致を問題にしている、主語は「が格」と考えてよい。B類の接続語句については、接続語句の前後の両方の省略に対しては文脈によるとしている。A類の接続語句についての、[接続語句の前後で主語が一致する]という規則は、今、上で述べた規則と対応している。このような制約は、3.2.3節で、ゼロ代名詞照応解析のアルゴリズムの中に組み込まれて利用されることになるであろう。

表1. 接続語句による主語の一致への制約 [26]

			接続語句
A類	前or後省略	接続語句の前後で 主語が一致する	ながら、て たり、つつ
	両方省略		
B類	前or後省略	接続語句の前後で 主語が一致しにくい	ても、ので けれど、ば から、のに と
	両方省略	文脈による	
C類	前or後省略	文脈による	が
	両方省略		

3. 複文におけるゼロ代名詞照応の解析

3.1 ゼロ代名詞照応解析の基本的な手順

ゼロ代名詞の解析は、明示的な代名詞の解析と違って、その存在を即時的に同定することができないところに一つの問題がある。すでに、2.1節で、その基本的な過程についてまとめているが、本節では、より一般的な解析モデルの構成を考えてみよう。複文の処理ということを考えながら、解析モデルの構成を進める [1,16]。まず、解析の基本的な手順をまとめる。

【ゼロ代名詞照応解析の基本的な手順】

- (1) [文章の一部] を読み、その [解析結果] に基づいて、ゼロ代名詞を同定する。
- (2) [文章の一部] の [解析結果] に基づいて、[制約条件] を設定する。
- (3) [制約条件] のもとで、[文脈情報] の中に照応要素および先行表現を探索する。

この手順の基本構成要素をもう少し具体的にみてみよう [16]。

・[文章の一部]

節：ゼロ代名詞の同定を認知する基本的な単位として、述語を一つ含む節を考える。

・[解析結果]

意味表現：動詞の格フレームに基づいて構成される節（文）の意味表現を考える。

動詞の格フレームの枠組み

(<動詞> <動詞意味属性>
 (<必須深層格> <意味属性> <表層格 (格助詞)>

 (<任意深層格> <意味属性> <表層格 (格助詞)>
 )

<動詞意味属性> : 動詞が持つ意味的な属性

・ [制約条件]

- (イ) 動詞の格フレームに含まれている情報から設定される制約
- ・ 格情報 (深層格, 意味属性, 表層格) からの制約
 - ・ 動詞意味属性からの制約 (中岩・池原 [19])
- (ロ) ゼロ代名詞を含む節 (文) に含まれている格情報以外の情報から設定される制約
- ・ 様相表現からの制約 (中岩・池原 [19])
 - ・ 待遇表現からの制約 (堂坂 [3])
 - ・ 視点に関わる受給表現からの制約 (久野 [11], 堂坂 [3])
 - ・ 情報の縄張りに関わる文末形式からの制約 (堂坂 [3])
 - ・ 接続語句からの制約 (吉本 [30], 田村・奥村 [26], 中岩・池原 [19])
 - ・ 意味および談話役割の間の関係からの制約 (中川・西澤 [20])
- (ハ) 文章の種類などの文章に関する情報から設定される制約
- ・ 一人称の物語における作者の役割から設定される制約

・ [文脈情報]

これまでに様々な文脈情報の保持と探索のしくみが提案されている [16, 22]。

- (イ) 最も素朴なしくみは, 2.1 節で考えた, [文章の一部] に出てきた順番に対象を並べたリストを保持して, 先頭から順番に探索するというしくみである。意味属性と格情報も加えて, (<指示対象> <先行表現> <意味属性> <格情報>) という 4 項組で対象を表現する。
- (ロ) さらに, 焦点, 中心, 顕著さ, 情報の新旧といった機能的な情報に基づいて, 文脈情報を構造化, 階層化, 詳細化して表現し, 探索するしくみが提案されている。

【ゼロ代名詞照応解析の基本的な構成要素】

- (1) [文章の一部] : 節 (文)。
- (2) [解析結果] : 節 (文) の意味表現。
- (2) [制約条件] : 格フレームからの制約を基礎として, 節 (文) から得られる制約。

- (3) [文脈情報] : 節(文)に現れる対象を出現順に並べたリスト。節(文)の解析が終わった時点で更新する。

上のような基本的な(素朴な)構成要素による基本的な手順で、例を一つ解析してみよう。

<27>あげはちょうが、みかんの木にとんできました。

ときどき、おなかの先をまげて、葉に、なにかつけています。

1: あげはちょうが、みかんの木にとんできました。

ゼロ代名詞を同定せず。

文脈情報を初期化する。

文脈リスト: (('あげはちょう', "あげはちょう", <動物>, <が: 動作主>)
('みかんの木', "みかんの木", <植物>, <に: 対象>))

2: ときどき、おなかの先をまげて、

ゼロ代名詞の同定: 「[φが] おなかの先をまげて」

制約条件 : [(φ<動物><が: 動作主>) [おなかの先をまげて]]

照応要素 : 'あげはちょう'

文脈情報を更新する。

文脈リスト: (('あげはちょう', "あげはちょう", <動物>, <が: 動作主>)
('おなかの先', "おなかの先", <動物の部分>, <を: 対象>))

3: 葉に、なにかつけています。

ゼロ代名詞の同定: 「[φが] 葉に、なにかつけています。」

制約条件 : [(φ<動物><が: 動作主>) [葉に、なにかつけています。]]

照応要素 : 'あげはちょう'

文脈情報を更新する。

文脈リスト: (('あげはちょう', "あげはちょう", <動物>, <が: 動作主>)
('葉', "葉", <植物部分>, <に: 対象>))

第2文は、「て形接続」節を含む複文である。これを二つの節に分けて解析している。処理中の節(文)のゼロ代名詞の照応要素の探索は前節(文)からの文脈リストに対して行われる。また、文脈リストの探索は先頭から始めるものとする。この例のような節(文)連続における主格ゼロ代名詞は、前節(文)の主格要素を継続して照応要素とすることが多く、前節の文脈リストを探索することで解析が成功する。

3.2 ゼロ代名詞照応解析へのセンター理論の適用と拡張

3.2.1 センター理論

センター理論(Centering Theory) (GJW[5,6])では、談話セグメント(Discourse Segment)

における発話 (Utterance) の中で参照される対象をセンター (centers) と呼ぶ。談話セグメントの各発話には、前向きセンター (forward looking centers) のリスト C_f と、(たかだか) 1個の後ろ向きセンター (backward looking center) C_b が割り当てられる。

(1) $C_f(DS, U_i)$: 前向きセンター (forward looking centers) のリスト

談話セグメント DS の発話 U_i で参照されている対象の集合である。 C_f の要素は、中心性の基準に基づいて順序付けられる。最も順位の高い要素を $C_p(DS, U_i)$ (preferred center) とする。 C_p は、 $C_b(DS, U_{i+1})$ となる可能性が最も高い。しかし、最終的に U_{i+1} で参照される $C_f(DS, U_i)$ の最も高い順位にある要素が実際の $C_b(DS, U_{i+1})$ である。

(2) $C_b(DS, U_i)$: backward looking center

U_i で参照される対象で、 $C_f(DS, U_{i-1})$ 中の最も高い順位にある要素である。

(3) C_f における対象の順位を主題性と文法的な機能を基準として次のように定める。

Topic > Subject > Object2 / Object > Others

(主題) (が格) (に格) (を格) (その他)

(次のような順位の設定も提案されている :

C_f Ranking by Grammatical Function (BFP [2])

Subject > Object2 > Object > Others

C_f Ranking for Japanese (WIC [27])

(Grammatical or zero) Topic > Empathy > Subject > Object2 > Object > Others)

(4) 次のような規則が設定される。

[R 1] $C_f(DS, U_{i-1})$ の任意の要素が U_i で代名詞によって参照されるならば、

$C_b(DS, U_i)$ も代名詞によって参照されるべきである。

[R 2] C_b の遷移について次のような優先順位が望ましい。

Continuing > Retaining > Smooth Shift > Rough Shift

それぞれの遷移は次のように設定される (BFP [2])。

	$C_b(U_i) = C_b(U_{i-1})$ or $C_b(U_{i-1}) = [? ; NO]$	$C_b(U_i) \neq C_b(U_{i-1})$
$C_b(U_i) = C_p(U_i)$	Continuing	Smooth Shift
$C_b(U_i) \neq C_p(U_i)$	Retaining	Rough Shift

簡略化のために、 C_f と C_b の引数から DS を省略して書いている。最も望ましい遷移は、Continuing であり、それは、すなわち、 C_b が継続すること、そして、 C_b と C_p が同じであることが望ましいということである。また、談話セグメント DS の先頭の発話 U_1 の backward-looking center すなわち $C_b(DS, U_1)$ は、参照する前発話がないので、[なし(?)]ということになる。参照する前発話はあるけれども、その前発話の C_f からの対象が一つも参照されていな

い発話における backward-looking center Cb も [なし] ということになる。この [なし] を [?] と区別して, [NO] とし, BFP [2] による遷移表に追加する。

ゼロ代名詞照応解析へのセンター理論の適用は, 文脈情報をセンター情報 Cf と Cb として, 次のような基本的な解析アルゴリズムへと導く [2,23,27]。

【センター理論を適用したゼロ代名詞照応解析アルゴリズム：A1】

- (1) U_i を読み, その解析結果に基づいて, ゼロ代名詞を同定する。
- (2) U_i の解析結果に基づいて, 制約条件を設定する。
- (3) $Cf(U_{i-1})$ の中に, 制約条件を満たす要素を探索する。
- (4) $Cb(U_i)$ と $Cf(U_i)$ を計算し, $Cf(U_i)$ の要素を順位付けする。
- (5) 複数の解が存在する場合は, 遷移の優先順位規則 [R2] に従って決定する。

このアルゴリズム A1 は, 文脈情報の参照の範囲を前文に限って, 照応関係の局所的な情報に基づく解析アルゴリズムを提案している。代名詞による照応は, その範囲を前文に限って行われるわけではないので, その点に問題を残しながらも, 発話単位を単文とする文の連続に対しては, 比較的有効に適用できるということが報告されている [2,6,8,27,28]。

3.2.2 複文へのセンター理論の適用

複文へのセンター理論の適用において, まず, 考えなければならないのは, 発話(Utterance)をどのような単位として設定するかということであるが, 2章での考察から明らかなように, 節を発話単位として設定するのが最も自然である。田村・奥村 [26] は, 複文を単文(節)+接続語句に分解して単文の列に変換し, その単文の列に対してセンター理論を適用して, ゼロ代名詞の補完を行う手法を提案している(2.3節)。その際, 接続助詞の機能を考慮した制約条件を用いて先行表現候補の絞り込みを行う。また, 連体節内のゼロ代名詞, 先行表現が文章中に明示されていないゼロ代名詞については考慮していない。Kameyama [9] も, 英語の複文へのセンター理論の適用の中で, 節を発話単位として, 線形のおよび階層的な節間の関係構造によって構成される複文の構造を考慮したセンター理論の適用を試みている。本節では, 田村・奥村 [26] による手法に従って, 複文を節の列に分解して, その節の列にセンター理論の適用を考える。具体例について, 前節でのアルゴリズム A1 を適用してみよう。アルゴリズム A1 の(1), (2), (3)の詳細な処理の記述は省略し, 照応要素と(4)と(5)の結果のみを示している。

<28>あげはちょうが, みかんの木に飛んできました。

[ϕ 1が] ときどきおなかの先をまげて, [ϕ 2が] 葉になにかつけています。

1: あげはちょうが, みかんの木にとんできました。

Cf: ('あげはちょう', 'みかんの木')

Cb: ?

- 2: [ϕ 1が] ときどきおなかの先をまげて、
 [ϕ 1の照応要素: 'あげはちょう']
 Cf: ('あげはちょう', 'おなかの先')
 Cb: 'あげはちょう' [Continuing]

- 3: [ϕ 2が] 葉になにかつけています。
 [ϕ 2: 'あげはちょう']
 Cf: ('あげはちょう', '葉')
 Cb: 'あげはちょう' [Continuing]

2, 3ともに、ゼロ代名詞の照応要素は、直前の節のCfの中に曖昧さなく見出すことができ、さらに、Cbの遷移は最も望ましい [Continuing] となっている。

次の例では、最初の文の文頭に主題表現があって、その指示対象が引き続く節および文の主格ゼロ代名詞の照応要素となっている。ここでは、複文中の主題表現は、すぐ右に隣接する節に属すると考えることにする。

<29>にほんざるは、えさを見つけると、 [ϕ 1が] [ϕ 2を] 早くたくさんほおぼって、
 [ϕ 3が] いそいでにげます。そして、 [ϕ 4が] 安全なばしょにいったから、 [ϕ 5が] [ϕ 6を] ゆっくりたべます。

- 1: にほんざるは、えさを見つけると、
 Cf: ('にほんざる', 'えさ')
 Cb: ?
- 2: [ϕ 1が] [ϕ 2を] 早くたくさんほおぼって、
 [ϕ 1: 'にほんざる'; ϕ 2: 'えさ']
 Cf: ('にほんざる', 'えさ')
 Cb: 'にほんざる' [Continuing]
- 3: [ϕ 3が] いそいでにげます。
 [ϕ 3: 'にほんざる']
 Cf: ('にほんざる')
 Cb: 'にほんざる' [Continuing]
- 4: そして、 [ϕ 4が] 安全なばしょにいったから、
 [ϕ 4: 'にほんざる']
 Cf: ('にほんざる', '安全なばしょ')
 Cb: 'にほんざる' [Continuing]
- 5: [ϕ 5が] [ϕ 6を] ゆっくりたべます。

[ϕ 5: 'にほんざる': ϕ 6: '*']

Cf: ('にほんざる', '*')

Cb: 'にほんざる' [Continuing]

上のような、前節の Cf のみを文脈情報とするアルゴリズム A1 は、最後の節の ϕ 6 の照応要素の解析に失敗する。センター理論にとっての文脈情報が前節の Cf に限られるところに問題がある。 ϕ 6 の照応要素は、3つ前の節2の Cf 中の 'えさ' である。田村・奥村 [26] の手法では、4文(節)前の Cf まで探索するという拡張を提案している。この拡張によれば、'えさ' を見いだすことができる。この例において、4文(節)前というのは、もとの文では前文の中の節である。一つの文が4つ以内の節で構成されるときは限らない。また、この処理では、複文中の主題表現の役割が十分に考慮されているとはいえない。もうひとつの例について、アルゴリズム A1 の適用を行ってみよう。物語文章への適用ということで、[書き手]への文章外照応の処理も仮定する。

<30> [ϕ 1が] 外へ出ようと思って、[ϕ 2が] げんかんの戸を開けると、かわいい子ねこがちょこんとすわっていました。わたしは、びっくりしましたが、[ϕ 3が] 寒そうにしているので、[ϕ 4が] [ϕ 5を] 家の中に入れてやりました。

1: [ϕ 1が] 外へ出ようと思って、

[ϕ 1: 'わたし': [書き手] への文章外照応]

Cf: ('わたし', '外')

Cb: ?

2: [ϕ 2が] げんかんの戸を開けると、

[ϕ 2: 'わたし']

Cf: ('わたし', 'げんかんの戸')

Cb: 'わたし' [Continuing]

3: かわいい子ねこがちょこんとすわっていました。

Cf: ('かわいい子ねこ')

Cb: NO [* Shift]

4: わたしは、びっくりしましたが、

Cf: ('わたし')

Cb: NO [* Shift]

5: [ϕ 3が] 寒そうにしているので、

[ϕ 3: 'わたし']

Cf: ('わたし')

Cb: 'わたし' [Continuing]

6 : [ϕ 4 が] [ϕ 5 を] 家の中に入れてやりました。

[ϕ 4 : 'わたし' ; ϕ 5 : '*']

Cf : ('わたし', '*')

Cb : 'わたし'

[Continuing]

文章外照応の処理をしなければ、この文章における 1, 2 のゼロ代名詞 ϕ 1, ϕ 2 は照応要素を見いだすことができない。残り 3 つのゼロ代名詞のうち、解析が成功しているのは、6 の ϕ 4 だけである。5 のゼロ代名詞 ϕ 3 と 6 のゼロ代名詞 ϕ 5 の照応要素は 'かわいい子ねこ' である。5 については、前節の Cf からは、'わたし' 以外、情報を得ることができない。また、この 'わたし' は、動詞の格フレームからの制約にも違反していない。しかし、これは間違いである。これを解決するためには、5 からの制約だけでなく前後の文脈情報と知識による制約が必須である。6 の ϕ 5 についても、5 の ϕ 3 が正しく解析されなければ、その照応要素を解析することができない。'*' は不定の照応要素を表わす。節 3 のように、2 つの文の連続 U_{i-1} , U_i で、 U_i はゼロ代名詞を一つも含まず、さらに U_{i-1} からの対象を一つも参照していないということは普通にあることである。つまり、これは、 U_i には Cb が無い (NO) ということである。ゼロ代名詞を含んだとしても、 U_i に Cb が無いということは不自然なことではない。

[* Shift] は [R 2] の遷移表では規定されない Shift である。この Shift についての考察が必要である。

3.2.3 複文へのセンター理論の拡張に関する従来の研究

本節では、複文におけるゼロ代名詞の解析が可能なようにセンター理論を拡張した、従来のいくつかの試みについて検討する。

① 田村・奥村 [26] による拡張については、前に言及した。複文を単文+接続語句に分解して単文の列に変換し、その単文の列に対してセンター理論を適用して、ゼロ代名詞の補完を行う。その際、接続語句の機能を考慮した制約条件を用いて先行表現候補の絞り込みを行う。連体節内のゼロ代名詞、先行表現が文章中に明示されていないゼロ代名詞は処理の対象としていない。さらに、前文の Cf だけでなく、4 文前までの Cf を保持して探索する。前節でも述べたように、このモデルで、前節のアルゴリズム A1 では解決できない問題のいくつかを解決することができる。しかし、複文を単文の列に分割して、線形に並べてしまうと、複文という単位が持っている特徴、複文と前後の文との間の関係などの有用な情報が失われてしまうように思われる。

② Kameyama [9] は、英語の複文における代名詞による照応の解析のために、センター理論の拡張を検討している。基本的な仮説として、複文はセンター理論における発話に対応する単位 (センター情報を更新する単位) に分割されるとする。そのような単位として、複文の構成要素である節を設定し、節を時制あり (tensed) と時制なし (untensed) に分類する。時制あ

りの節について、複文の基本的な構造である線形的な構造と階層的な構造に対応して、二つのセンター処理, Sequential intrasentential centering と Hierarchical intrasentential centering を設定する。それぞれについてコーパスからの具体例に基づいて、いくつかの処理仮説を提案している。複文の構造を考慮に入れること、複文内の節間の接続関係についても考慮しなければならないということなど、複文へのセンター理論の拡張にとって基本的な提案が具体例による分析とともに示されている。

③高田・土居 [25] によるセンターリストモデルでは、[代名詞・ゼロ代名詞として文に現れた実体は、通常の定名詞句として現れた実体よりも、次の文で代名詞・ゼロ代名詞になる可能性が高い] という認識に基づいて、次のような文脈保持のためのリストが構成される。

・センターリスト (CL) : ゼロ代名詞または代名詞になった実体を保持する。

・可能センターリスト (PoCL) : 陽に表現されたその他の実体を保持する。

各リスト内で実体はさらに順序付けられる。

Top > Sub > Obj > Obj 2 > Others (主節)

> Sub / Obj / Obj 2 (従属節) > Others (従属節)

この順序付けの中で、複文を構成する副節(従属節)の中の対象への順序付けも行われている。しかし、副節(従属節)は一つとは限らず、副節が2つ以上あって、ゼロ代名詞も2つ以上あるような場合に、副節の中での順位をどのように処理するのか明らかでない。CL, PoCLともに、現在の文から構成され、それが次の文の処理のためにわたされる。つまり、センター理論における Cf を CL と PoCL に分割したと考えられる。そして、CLの方がPoCLより先に探索される。 Cf = CL ∪ PoCL (CL > PoCL)。

さらに、先行詞が必ずしも前文にあるとは限らないので、過去センターリスト (PaCL) と名詞リスト (NL) を付け加える。これらのリストに入っている対象は、現在の文からの距離を考慮して順序付けられる。これによって、文脈中に現われたすべての対象を保持するという仕組みは構成されたが、複文の構造や副節間の接続関係などを十分に考慮しているとは言い難い。

④ Strube & Hahn [23] による機能的センタリング (Functional Centering) では、情報状態 (Information Status : 情報の新旧) に基づいて、Cf 中の対象を順序付けるモデルを提案している。聞き手(読み手)にとって古い対象は新しい対象よりも順位が先にあると考え、発話 U_i における対象を情報の新旧に応じて、次のように分類する【基本的な分類】。

DE : the set of discourse entities in U_i

E : the set of evoked discourse entities (既出の談話対象) in U_i

U : the set of unused discourse entities (既出ではないが、既知の談話対象) in U_i

OLD := E ∪ U (旧対象の集合)

NEW := DE - OLD (新対象の集合)

そして、Cf 中の対象の順位を次のように設定する【基本的な順位付け制約】。

1. If $x \in \text{OLD}$ and $y \in \text{NEW}$, then $x \ll y$.
2. If $x, y \in \text{OLD}$ or $x, y \in \text{NEW}$, then $x \ll y$, if $\text{pos}x < \text{pos}y$.
3. If (1) or (2) do not apply, then x and y are unordered with respect to the Cf ranking.

($x \ll y$: x が y より順位が先; $\text{pos}x < \text{pos}y$: $\text{pos}x$ が $\text{pos}y$ より位置が前)

さらに、上述の対象の基本的な分類と順位付け制約を、推論による照応の処理も考えて拡張したモデルを構成している。複文への対応としては、Kameyama [9]に従って、発話単位を節のレベルで捉えるという考え方をとっている。しかし、複文の構造や副節間の接続関係についての考慮がどのようになされるかについての明白な記述は行われていない。

3.3 複文の構造を考慮したセンター情報の構成

本節では、これまでの考察を踏まえて、日本語複文の構造を考慮に入れたセンター情報の構成とそれに基づく照応要素の探索アルゴリズムについて考える。複文は、一般に、複数の節から構成され、線形的、あるいは階層的な構造を持っている。従って、節を発話単位として計算される Cf を、そのような複文の構造を考慮に入れて保持し利用するしくみを考えるのが自然であると思われる。さらに、複文においては、2章の多くの例でみてきたように、主題表現の指示対象である主題要素が特別な役割を担っており、この主題要素の役割を考慮に入れたしくみを考えることも基本である。本節でのセンター情報の構成においては、主題要素を一つの独立したセンターとし、特別な優先順位を与えることにする。

文章は、文の並び $S_0 \cdots S_{i-1} S_i S_{i+1} \cdots S_n$ から構成されるものとする。ここで、解析アルゴリズムが対象とする複文の副節は、副詞節と並列節だけに限るものとする。その他の節についての考慮は今後の課題である。節を U_j とする。Cf は節ごとに計算する。新しいセンター C_t と C_w 、および C_t と Cf のリスト $C_f s$ を考える。 C_t は、主題要素である。 C_w は、物語文章における作者(書き手)に対応する。文 S_i に対して、 C_t の初期値は空とする。文 S_i の中で、主題表現の出現とともに更新され、文の処理が終わった時に、最近の C_t が $C_f s$ の中に保存される。

・ S_i が単文の時は、 $U_i = S_i$ で、 $C_f s(S_i) = (C_t(S_i) \text{ Cf}(U_i))$ 。

・ S_i が複文の時は、 S_i を構成する副節を $U_{i1} \cdots U_{in}$ とすると、

$C_f s(S_i) = (C_t(S_i) \text{ Cf}(U_{in}) \text{ Cf}(U_{in-1}) \cdots \text{ Cf}(U_{i1}))$ 。

$C_f s$ の要素はトップから、 $t, 1, 2, \dots$ として参照される。

今、複文 S_i 中の節 U_{ij} においてゼロ代名詞を同定したとする。以上の設定で、センター情報の探索アルゴリズムを次のように構成する。

【センター情報の探索アルゴリズム】

- (1) Ct(Si)を探索する。
- (2) Cf(Uij-1), Cf(Uij-2), ..., Cf(Ui 1)を探索する。
- (3) Cfs(Si-1)を探索する。
- (4) Cwを探索する。
- (5) Cfs(Si-2), Cfs(Si-3), ... を探索する。

Ctを保持して、最優先で探索する。次に、まず、Uijが含まれている複文Si中の、Uijより前の節のCfを探索し、次に、Uijが含まれている複文の前文Si-1を構成する節のCfとCtのリストCfsを探索する。このとき、最初に前文の最近のCtが探索されることになる。最近のCtのみとしているが、これは検討の余地がある。このような探索の戦略は、複文の構造に依存した自然な戦略であると考えられる。物語文章という特徴に依存して、Cwを保持して探索する。Cw='わたし'である。最後に、2文以上前のCfとCtのリストCfsを探索する。この探索アルゴリズムを組み込んだゼロ代名詞照応解析のアルゴリズムを次のように構成する。

【ゼロ代名詞照応解析アルゴリズム：A2】

- (1) 主題表現を読んだら、Ctを更新する。
- (2) Uiを読み、その解析結果に基づいて、ゼロ代名詞を同定する。
- (3) Uiの解析結果に基づいて、探索のための制約条件を設定する。
- (4) 制約条件を満たす要素を探索アルゴリズムにより探索する。
- (5) Cf(Ui)を計算し、Cf(Ui)の要素を順位付けする。
- (6) 文の終わりであれば、Cfsを構成する。

具体例について、このアルゴリズムA2によりゼロ代名詞の解析を行ってみよう。Cf(Ui)の要素の順序付けは、主格を先頭に文法的な機能に基づく順序とする。アルゴリズムA1との比較を行うために、一つ前の節に対してCbの計算も行う。

<31> 1. にほんざるは、[φ1が]えさを見つけると、[φ2が][φ3を]早くたくさんほおばって、[φ4が]いそいでにげます。

2. そして、[φ5が]安全なばしょにいったから、[φ6が][φ7を]ゆっくりたべます。

1. にほんざるは、 [Ct='にほんざる']

[φ1が] えさを見つけると、

Cf= ('にほんざる', 'えさ') : φ1=Ct

Cb=?

[φ2が][φ3を] 早くたくさんほおばって、

Cf= ('にほんざる', 'えさ') : φ2=Ct, φ3=Cf(-1:2)

Cb='にほんざる' [Continuing]
 [φ4が] いそいでにげます。
 Cf= ('にほんざる') : φ4=Ct
 Cb='にほんざる' [Continuing]
 Cfs= (('にほんざる') ('にほんざる') ('にほんざる', 'えさ')
 ('にほんざる', 'えさ'))

2. そして, [Ct=' ']
 [φ5が] 安全なばしょにいったから,
 Cf= ('にほんざる', '安全なばしょ') : φ5=Cfs(-1:t)
 Cb='にほんざる' [Continuing]
 [φ6が] [φ7を] ゆっくりたべます。
 Cf= ('にほんざる', 'えさ') : φ6=Cf(-1:1), φ7=Cfs(-1:(2:2))
 Cb='にほんざる' [Continuing]
 Cfs= (() ('にほんざる', 'えさ') ('にほんざる', '安全なばしょ'))

ゼロ代名詞の照応要素に対応するセンター情報と遷移をまとめると次のようになる。

1 : 1 : [Ct] φ1=Ct
 : 2 : φ2=Ct, φ3=Cf(-1:2) [Continuing]
 : 3 : φ4=Ct [Continuing]
 2 : 1 : φ5=Cfs(-1:t) [Continuing]
 : 2 : φ6=Cf(-1:1), φ7=Cfs(-1:(2:2)) [Continuing]

一番速くて、Cfs(-1:(2:2))である。前文(-1と表わす)のCfsの2番目のCfの2番目((2:2))の要素として解析される。そして、この例では、すべての遷移が[Continuing]となっている。

<32> 1. 冬休みに入ってまもない日の朝でした。

2. [φ1が] 外へ出ようと思って、[φ2が] げんかんの戸を開けると、かわいい子ねこがちょこんとすわっていました。

3. わたしは、[φ3が] びっくりしましたが、[φ4が] 寒そうにしているので、[φ5が] 家の中に入れてやりました。

1. 冬休みに入ってまもない日の朝でした。 [Ct=' ']
 Cf= ('冬休みに入って間もない日の朝')
 Cb=?
 Cfs= (() ('冬休みに入って間もない日の朝'))

2. [ϕ 1が] 外へ出ようと思って、 [Ct=' ']
 Cf= ('わたし', '外') : ϕ 1=Cw (= 'わたし')
 Cb=NO [* Shift]
- [ϕ 2が] げんかんの戸を開けると、
 Cf= ('わたし', 'げんかんの戸') : ϕ 2=Cf(-1:1)
 Cb='わたし' [Continuing]
 かわいい子ねこがちょこんとすわっていました。
 Cf= ('かわいい子ねこ')
 Cb=NO [* Shift]
 Cfs= (()) ('かわいい子ねこ') ('わたし', 'げんかんの戸') ('わたし', '外')
3. わたしは、 [Ct='わたし']
 [ϕ 3が] びっくりしましたが、
 Cf= ('わたし') : ϕ 3=Ct
 Cb=NO [* Shift]
- [ϕ 4が] 寒そうにしているので、
 Cf= ('わたし') : ϕ 4=Ct
 Cb='わたし' [Continuing]×
 Cf= ('かわいい子ねこ') : ϕ 4=Cfs(-1:(1:1))
 Cb=NO [* Shift] ○
- [ϕ 5が] [ϕ 6を] 家の中に入れてやりました。
 Cf= ('わたし', 'かわいい子ねこ') : ϕ 5=Ct, ϕ 6=Cf(-1:1)
 Cb='かわいい子ねこ' [RoughShift]
 Cfs= (()) ('わたし', 'かわいい子ねこ') ('かわいい子ねこ') ('わたし')

この例で、問題は、 ϕ 4の解析である。アルゴリズムA1におけると同様に、まず ϕ 4の照応要素は、'わたし'となる。ここで、文脈情報と知識の利用に基づく制約を考慮すると、これは間違いであると判定される。間違いであることを、×で示している。 ϕ 4の照応要素として、'かわいい子ねこ'が正しい。正しいことを○で示している。アルゴリズムA2では、正しい照応要素をCfs(-1)の中に探索することができる。そして、文脈情報と知識の利用に基づく意味的制約を満たす照応要素を決定することができる。

上の例について、アルゴリズムA2で正しい解が得られた場合に、各節のゼロ代名詞の照応要素がどこから探索されたかをまとめると、次のようになる。Cfs(-1)の範囲までにあるということがわかる。[* Shift]は、Cb=NOに対応する、センター理論では規定されていない遷移である。

- 1 : 1 :
 2 : 1 : $\phi 1 = Cw$ [\ast Shift]
 : 2 : $\phi 2 = Cf(-1 : 1)$ [Continuing]
 : 3 : [\ast Shift]
 3 : 1 : [Ct] $\phi 3 = Ct$ [\ast Shift]
 : 2 : $\phi 4 = Cfs(-1 : (1 : 1))$ [\ast Shift]
 : 3 : $\phi 5 = Ct, \phi 6 = Cf(-1 : 1)$ [RoughShift]

<33> 1. モユちゃんは、 $[\phi 1$ が] 友子にピシャピシャ顔をたたかれながら、 $[\phi 2$ が] きよとんと友子を見上げていました。

2. $[\phi 3$ が] それを見ると、友子は、 $[\phi 4$ が] なみだをぼろぼろこぼしながら、 $[\phi 5$ が] モユちゃんにほおずりました。

1. モユちゃんは、 [Ct='モユちゃん']
 $[\phi 1$ が] 友子にピシャピシャ顔をたたかれながら、
 Cf= ('モユちゃん', '友子', '顔') : $\phi 1 = Ct$
 Cb= ?
 $[\phi 2$ が] きよとんと友子を見上げていました。
 Cf= ('モユちゃん', '友子') : $\phi 2 = Ct$
 Cb= 'モユちゃん' [Continuing]
 Cfs= (('モユちゃん') ('モユちゃん', '友子') ('モユちゃん', '友子', '顔'))
2. $[\phi 3$ が] それを見ると、 [Ct=' ']
 Cf= ('モユちゃん', 'それ') : $\phi 3 = Cfs(-1 : t)$
 Cb= 'モユちゃん' [Continuing] ×
 Cf= ('友子', 'それ') : $\phi 3 = Cfs(-1 : (1 : 2))$
 Cb= '友子' [SmoothShift] ○
 友子は、 [Ct='友子']
 $[\phi 4$ が] なみだをぼろぼろこぼしながら、
 Cf= ('友子', 'なみだ') : $\phi 4 = Ct$
 Cb= '友子' [Continuing]
 $[\phi 5$ が] モユちゃんにほおずりました。
 Cf= ('友子', 'モユちゃん') : $\phi 5 = Ct$
 Cb= '友子' [Continuing]
 Cfs= (('友子') ('友子', 'モユちゃん') ('友子', 'なみだ') ('友子', 'それ'))

この例で、問題は、 $\phi 3$ の解析である。アルゴリズム A1においては、 $\phi 4$ の照応要素の候補として、'モユちゃん'と'友子'が得られ、次に、遷移に関する優先順位規則 [R2]によって、'モユちゃん'が選好される。しかし、これは間違いである。 $\phi 4$ の照応要素としては、'友子'が正しい。アルゴリズム A1で探索可能な要素の中に複数の候補があるとき、遷移に関する優先順位規則 [R2]が正しく機能しない例である。この例では、前節までのセンター情報だけでは、 $\phi 3$ の照応要素を正しく解析することができない。正解を得るためには、前方の文脈だけでなく、後続する文脈からの情報も必要であると思われる。これは、同時にアルゴリズム A2における問題でもある。

4. おわりに

日本語文章の複文におけるゼロ代名詞照応の解析のための枠組みについて、従来の研究を参考にしながら、基礎的な考察を行った。複文は、一般に複数の副節を含んで、線形的あるいは階層的な、様々な構造を持って存在しており、それらの構造を十分考慮に入れた、ゼロ代名詞照応の解析アルゴリズムを構成しなければならない。文脈情報の保持と探索のしくみとして、センター理論におけるセンター情報とその処理機構を拡張したしくみを構成して、具体的な文章例への適用を試みた。しかし、まだ解析の対象とする文章例も少なく、副節を副詞節と並列節のみに限っており、今後の多くの考察が残されている。特に、ゼロ代名詞を含む節(文)を超えた文脈からの情報や知識に基づく制約についての考察は重要である。また、前節の最後の例のような、前方からのセンター情報の利用だけでは正解を得ることのできないゼロ代名詞照応の解析のためのしくみについての考察も今後の課題である。

参考文献

- [1] 阿部純一・桃内佳雄・金子康朗・李光五：人間の言語情報処理－言語理解の認知科学，サイエンス社，1994.
- [2] Brennan, S.E., Friedman, M.W. & Pollard, C.J.: A centering approach to pronouns Proc. of the 25th annual meeting of the ACL, pp. 155-162, 1987.
- [3] 堂坂浩二：語用論的条件の解釈に基づく日本語ゼロ代名詞の指示対象同定，情報処理学会論文誌，Vol. 35, No. 5, pp. 768-778, 1994.
- [4] 藤澤伸二・増山繁・内藤昭三：日本語文章における照応・省略現象の基本検討，情報処理学会論文誌，Vol. 34, No. 9, pp. 1909-1918, 1993.
- [5] Grosz, B.J., Joshi, A.K. & Weinstein S.: Providing a unified account of definite noun phrases in discourse, Proc. of 21st Annual Meeting of the ACL, pp. 44-50, 1983.
- [6] B.J. Grosz, A.K. Joshi and S. Weinstein: Centering: A Framework for Modeling the Local Coherence of Discourse, Computational Linguistics, Vol. 21, No. 2, pp. 203-225, 1995.

- [7] 長谷川信子：省略された代名詞の解釈，日本語学，Vol. 14, No. 4, pp. 27-34, 1995.
- [8] Iida, M.: Discourse Coherence and Shifting Centers in Japanese Texts, In M.A. Walker, A.K. Joshi & E.F. Prince(eds), Centering Theory in Discourse, Oxford University Press, 1998.
- [9] Kameyama, M.: Intrasentential Centering: A Case Study, In M.A. Walker, A.K. Joshi and E.F. Prince(eds), Centering Theory in Discourse, Oxford University Press, 1998.
- [10] Kameyama, M.: Infeasible Semantics and Defeasible Pragmatics, In M. Kanazawa, C. Pinon & H. de Swart(eds), Quantifiers, Deduction, and Context, CSLI Publications, 1996.
- [11] 久野暲：談話の文法，大修館書店，1978.
- [12] 益岡隆志・田窪行則：基礎日本語文法-改訂版-，くろしお出版，1992.
- [13] 三上章：文法小論集，くろしお出版，1970.
- [14] 南不二夫：現代日本語の構造，大修館書店，1974.
- [15] 南不二夫：現代日本語文法の輪郭，大修館書店，1993.
- [16] 桃内佳雄，日本語文章におけるゼロ代名詞による照応の解析について，北海学園大学工学部研究報告，第24号，pp. 131-147, 1997.
- [17] 森岡健二：伝達論からみた省略，言語生活，No. 339, pp. 18-31, 1980.
- [18] 村田真樹・長尾眞：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，自然言語処理，Vol. 4, No. 1, pp. 87-110, 1996.
- [19] 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，自然言語処理，Vol. 3, No. 4, pp. 49-63, 1996.
- [20] 中川裕志・西澤信一郎：日本語の理由-行為の順接複文におけるゼロ代名詞照応，情報処理学会論文誌，Vol. 35, No. 10, pp. 2038-2045, 1994.
- [21] 野田尚史，「は」と「が」，くろしお出版，1996.
- [22] Sidner, C.L.: Focussing in the comprehension of definite anaphora, In M. Brady & R.C. Berwick(eds), Computational models of discourse, The MIT Press, 1983.
- [23] Strube, M. & Hahn, U.: Functional Centering—Grounding Referential Coherence in Information Structure, Computational Linguistics, Vol. 25, No. 3, pp. 309-344, 1999.
- [24] Suri, L.Z. & McCoy, K.E.: RATFT/RAPR and Centering: A Comparison and Discussion of Problems Related to Processing Complex Sentences, Computational Linguistics, Vol. 20, No. 2, pp. 301-317, 1994.
- [25] Takada, S. & Doi, N.: An Extended Centering Mechanism for Interpreting Pronouns and Zero-pronouns, IEICE Trans. INF & SYST., Vol. E78-D, No. 1, pp. 58-67, 1995.
- [26] 田村浩二・奥村学：センター理論による日本語談話の省略解析，情報処理学会研究報告，自然言語処理 107-12, pp. 91-96, 1995.
- [27] Walker, M., Iida, M. & Cote, S.: Japanese Discourse and the Process of Centering, Computational Linguistics, Vol. 20, No. 2, pp. 193-231, 1995.
- [28] Walker, M., Joshi, A. and Prince, E.: Centering in Naturally Occurring Discourse: An Overview, In M.A. Walker, A.K. Joshi & E.F. Prince(eds), Centering Theory in Discourse, Oxford University Press, 1998.
- [29] 山梨正明：推論と照応，くろしお出版，1992.
- [30] 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，情報処理学会研究報告，自然言語処理 56-7, pp. 1-8, 1986.