

タイトル	日本語ゼロ代名詞解析における制約の適用について
著者	桃内, 佳雄; 柴田, 更紗
引用	北海学園大学工学部研究報告, 34: 111-127
発行日	2007-02-20

日本語ゼロ代名詞解析における制約の適用について

桃内 佳雄*・柴田 更紗*

Constraints in Interpretation of Japanese Zero Pronoun

Yoshio MOMOUCHI* and Sarasa SHIBATA*

あらまし

日本語文章におけるゼロ代名詞解析のための基本的なモデルを構成し、ゼロ代名詞の指示対象候補を保持するしくみと指示対象候補の妥当性の検証あるいは最適性の評価のための制約の適用に関する基礎的な考察を進める。指示対象候補への制約の適用というしくみは、ゼロ代名詞解析における知識と文脈情報の統合的な処理方式の構成にとって有効なしくみであると考えられる。

1. はじめに

日本語文章におけるゼロ代名詞解析については、様々な視点からの考察が進められており、いくつかの解析モデルが提案されている¹⁻⁶⁾。本報告では、亀山によるCENTERモデル^{2,3)}を基礎として三つのモデルを段階的に構成して考察を進める。それらのモデルの構成におけるゼロ代名詞解析の基本的な処理過程は次のようにまとめられる。

- (1) ゼロ代名詞の指示対象候補の保持：中心性に基づく構成と保持。
- (2) ゼロ代名詞の同定：動詞・名詞の意味構造に基づく同定。
- (3) ゼロ代名詞の指示対象の決定：制約に基づく決定。

文脈中の指示対象の保持のしくみを考えるとき、亀山によるCENTERモデルは、指示対象候補として一個の対象（CENTER）を保持するモデルである。これを最小の保持数として、文脈中に出現したすべての対象を保持するモデルまで、段階的、構造的な保持のしくみのモデルを考えることができる。本報告で構成するCENTERモデルを拡張したCENTER++モデルは、トピ

* 北海学園大学大学院工学研究科電子情報工学専攻

* Division of Electronics and Information Engineering, Graduate School of Engineering, Hokkai-Gakuen University

ック（主題）を特別に取り立てるモデルである。成山によるSRLモデル^{4,5,6)}は、文（節）に出現するすべての明示的な対象を保持するモデルである。石田・藤井⁶⁾はSRLモデルを拡張している。本報告では、SRLモデルと石田・藤井による拡張を参照してORLモデルを構成する。

最適性理論を適用した代名詞解析のモデル^{7,8,9,10)}が提案されているが、これらは、原理的には文脈中に現れるすべての対象を代名詞の可能な指示対象候補とするモデルであると考えることができる。最適性理論に基づく代名詞解析においては、代名詞の指示対象の解決において、いくつかの制約を設定して、制約を満たすことにおいて最適な（調和的な）候補を代名詞の指示対象として決定する。本報告では、指示対象の段階的、構造的な保持のしくみと最適性理論に基づくモデルにおける制約による指示対象の解決のしくみを組み合わせたゼロ代名詞解析のアルゴリズムの構成について基礎的な考察を進める。ゼロ代名詞の指示対象の解決における制約の適用のしくみは、いくつかの問題点を含んではいるが、ゼロ代名詞解析における知識と文脈情報の統合的な処理方式の構成にとって有効なしくみであると考えられる。

2. ゼロ代名詞解析の基本的な処理

ゼロ代名詞は、もし表現するとすれば名詞（句）に対応する表現の省略形である。例として、次の文章¹⁴⁾におけるゼロ代名詞について考えてみよう。

<1>①あげはちょうが、みかんの木に飛んできました。

②ときどき、おなかの先をまげて、葉になにかつけています。

③いったいなにをしているのでしょうか。

④たまごをうんでいるのです。

この文章は、4文から構成されている。②、③、④で、主格（が格）および属格（の格）に対応する表現が省略されていると考えることができる。ゼロ代名詞を記号 ϕ 、主格（が格）であることを“が”、属格（の格）であることを“の”を用いて表すことにすれば、省略現象を次のように表現することができる。

<1>①あげはちょうが、みかんの木に飛んできました。

②ときどき、 $[\phi_1が][\phi_2の]$ おなかの先をまげて、 $[\phi_3が][\phi_4の]$ 葉になにかつけています。

③いったい $[\phi_5が]$ なにをしているのでしょうか。

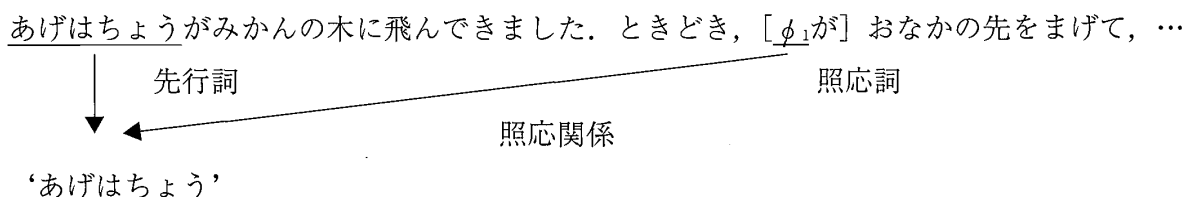
④ $[\phi_6が]$ たまごをうんでいるのです。

ゼロ代名詞を表層表現として復元すると次のようになる。

$\phi_1, \phi_2, \phi_3, \phi_5, \phi_6$ ：“あげはちょう”（①中の表現）

ϕ_4 ：“みかんの木”（①中の表現）

例えば、ゼロ代名詞 ϕ_1 と“あげはちょう”の間の関係を次のような関係として捉える。



照応要素（先行詞“あげはちょう”とゼロ代名詞 ϕ_1 の指示対象）

文脈を参照するゼロ代名詞を照応詞，参照される文脈中の表現“あげはちょう”を先行詞と呼ぶ．先行詞“あげはちょう”の指示対象を‘あげはちょう’とする．ゼロ代名詞 ϕ_1 の指示対象は，①の“あげはちょう”の指示対象‘あげはちょう’と同じであると考え．また，ゼロ代名詞の指示対象を照応要素と呼ぶ．上の例のような照応詞と先行詞／照応要素の関係性を照応関係と呼ぶ．

ゼロ代名詞解析のための基本的な処理は次の三つである．

(1) ゼロ代名詞の指示対象候補の保持

ゼロ代名詞の照応要素の候補となる対象を保持する．保持の構造をリストとすると，その要素間の順序関係（構造）をどのように考えるかが問題となる．

(2) ゼロ代名詞の同定

基本的には，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から同定する．動詞の意味フレーム（格フレーム）情報を知識として持ち，それを利用する．また，名詞が意味する概念にとっての必須修飾要素が欠けているということからも同定する．名詞の意味フレーム情報を知識として持ち，それを利用する．

(3) ゼロ代名詞の指示対象の決定

ゼロ代名詞の指示対象を，文脈または状況から生成され，保持されている指示対象候補のリストの中から，構文的・意味的・語用論的制約の適用というしくみにより決定する．

文章の解析を進めながら，文脈や状況の中から生成されてくる照応要素の候補となる対象をどのように構成，保持していくか，ゼロ代名詞をどのような手がかりにより同定するか，同定されたゼロ代名詞に対する照応要素を構文的・意味的・語用論的な制約の適用によりどのように決定するかが基本的な問題である．

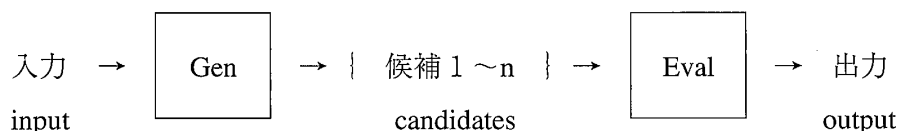
3. 最適性理論によるゼロ代名詞解析

3.1 最適性理論の枠組み

最適性理論（Optimality Theory）は，A.PrinceとP.Smolenskyによって発表された草稿から広く知られるようになった言語の理論であり，著書¹²⁾としてまとめられている．主に音韻論研究において広く応用されている．しかし，構文論や意味論における応用も進められ，代名詞の解析への応用もいくつか試みられている^{7,8,9,10,11)}．ここでは，最適性理論とそのゼロ代名詞解析へ

の応用の基本的な考え方について述べる。

最適性理論は、与えられた入力に対する最適な、調和のとれた解の候補を、制約をチェックすることにより求め、出力するしくみである。その基本的な構成は次のような図によって表現することができる¹³⁾。



・ Gen (Generator)

入力に対して、可能な候補を生成する。理論としては無限に多数の候補を生成する。

$$\text{Gen}(\text{input}) = \{ \text{candidate}_1, \text{candidate}_2, \dots, \text{candidate}_n \}$$

・ Eval (Evaluator)

ランク付けされた複数の制約を持つ。制約チェックの結果として、最適な解を出力する。

$$\text{Eval}(\text{candidate}_1, \text{candidate}_2, \dots, \text{candidate}_n) = \text{output}$$

制約をどのように設定し、ランク付けするかが、最適性理論における重要な問題である。

最適性理論のゼロ代名詞解析への応用においては、上の構成を次のような対応として考える。

入力：ゼロ代名詞（ゼロ代名詞は同定されているものとする）

候補：ゼロ代名詞の可能な指示対象候補

出力：ゼロ代名詞の最適な指示対象候補

制約：ゼロ代名詞解決のための制約

例えば、重要な制約として、選択制限がある。

選択制限 [指示対象の意味特徴は、ゼロ代名詞に対応する動詞の格フレーム要素の意味特徴と一致しなければならない.]

制約は、一般に複数個あって、優先順位によるランク付けがなされる。

3.2 ゼロ代名詞解析への最適性理論の適用

ゼロ代名詞解析のために利用される制約は多くあるが、Bouma et al.¹⁴⁾は、異なるモジュールからの情報という視点から、制約を次のように分類している。

①構文的な情報 (syntactic information)

agreement (一致)

configurational binding restrictions (束縛制限)

②意味的な情報 (semantic information)

selectional properties (動詞の選択的な性質：選択制限)

③語用論的な情報 (pragmatic information)

salience (指示対象の中心性, 顕著性)

④世界知識 (world knowledge)

これらの制約は常に同じ強さで作用するわけではない。制約には強さのランクが設定される。

ゼロ代名詞解析における制約の基本的な処理過程を見るために、次のような制約を設定して、具体的な日本語のゼロ代名詞解析の例について考えてみよう。これらの制約の妥当性については再考の余地がある。

SELECT: 先行詞の意味特徴はゼロ代名詞に対応する格フレーム要素の意味特徴と一致する。

DISJOINT: 述語の共出現項は排他的である。

SUBJECT: 主格ゼロ代名詞は前文の主格先行詞を選好する。

DISTANCE: ゼロ代名詞と先行詞の距離はより短いほうがよい。

制約のランクを優先順位として設定する。この優先順位の妥当性についても検討の余地がある。

SELE (CT) > DISJ (OINT) > SUBJ (ECT) > DIST (ANCE)

ゼロ代名詞解析処理は、文章を左から右に読み進みながら漸進的に行われるものとする。

2章での小学校国語教科書中のある文章¹⁴⁾の最初の2文を例として考えてみよう。

< 1 > ①あげはちょうが、みかんの木に飛んできました。

②ときどきおなかのさきを [ϕ_1 が] まげて、葉に、なにかつけています。

述語「まげて」まで読み進んだところでゼロ代名詞が同定される。それをゼロ代名詞 ϕ_1 とする。次のような照応要素の候補と制約から構成されるテーブルによって制約のチェックを行う。このテーブルを制約チェックテーブルと呼ぶことにする。制約を満たしていない場合に「×」を書く、協調的な比較評価が必要な場合に、満たしている場合のところに「○」を書く。

ϕ_1	制 約			
	SELE	DISJ	SUBJ	DIST
あげはちょう ▽			○	...
みかんのき	×		×	..
おなかのさき	×	×	×	.

ここでは、候補としては、文章中に出現している対象のすべてを出現の順に生成している。

Gen (ϕ_1) = { あげはちょう, みかんの木, おなかの先 }

これらの候補を4つの制約によりチェックすると、上位の制約をすべてクリアしている「あげはちょう」が最適な候補として選ばれて出力されることになる。DISTの評価は、決定的なものではないと考えている。文脈の中で指示対象候補が出現順の何個先にあるかを「・」の個数で示している。最適な指示対象候補には、▽印を付与している。

Eval [SELECT, DISJOINT, SUBJECT, DISTANCE]({ candidates }) = 「あげはちょう」

これは妥当な解である。

最適性理論に基づくゼロ代名詞解析を処理手順として構成すると次のようになる¹⁰⁾。

[最適性理論に基づくゼロ代名詞解析の処理手順]

- ①Create：与えられた入力I (ϕ ：ゼロ代名詞) に対して可能な指示対象候補の集合を生成する。この集合をGen (I) とする。その要素は候補として参照される。
- ②Check：各候補について、どの制約を満たすか、あるいは満たさないかを調べる。
- ③Calculate：各候補の調和ランキングを計算する。すなわち、各候補が制約をどのように満たすかを計算する。ある制約を満たすことが他の制約を満たすことよりもより重要であるという制約の間の優先順位を考慮に入れながら。
- ④Output：最も調和的な候補（最適な候補）を選択し、出力する。

Gen (I) をどのように生成するかが問題である。上の例< 1 >では、文脈に現れた可能なすべての対象を生成している。文脈が小さいので、3個の対象ですんでいるが、談話が展開していくにつれて文脈に現れる対象は増加していく。これをどのように処理するか、あるいは制御するかが問題である。最初の出発点としては、上の例のように文脈中に現れたすべての対象を可能な候補とするという方略となるであろう。また、調和ランキングの計算を具体的にどのように行うかも問題である。本報告では、上例のように、また、以下の解析においても、『制約の違反が、ランクとしてより低く、また数としてより少ないこと』を基準として考えることにする。

3.3 ゼロ代名詞解析のもう一つの例

もう一つの例について検討する。ただし、以下に例として示されている文章¹⁴⁾はより大きな文章の中の一部であり、ここでの可能な指示対象の候補は、示されている文章の中のみから生成されている。制約のチェックということに主眼をおいて考える。3.2で設定した制約に加えて、新しい制約の設定についても考える。

< 2 >①外へ出ようと思って、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。

- ②わたしは、 $[\phi_1が]$ びっくりしましたが、
- ③ $[\phi_2が]$ 寒そうにしているので、
- ④ $[\phi_3が][\phi_4を]$ 家の中に入れてやりました。

② ϕ_1 : わたし

ϕ_1	SELE	DISJ	CTOP	SUBJ	DIST
外	×			
げんかんの戸	×			
かわいい子ねこ			×		..
わたし ▽			○		.

CTOP (CTOPIC): 副節の「が格ゼロ代名詞」は主題が設定されていれば主題を照応する.

SELE: 述語「びっくりする」の主体の意味特徴は「animate (有生)」である.

③ ϕ_2 : かわいい子ねこ

ϕ_2	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
外		×			
げんかんの戸		×			
かわいい子ねこ ▽	○			×		..
わたし	×			○		.

DIFF (DIFFER): 述語「~そうにしている」の主体は発話者ではない第3者である.

SELE: 述語「寒そうにしている」の主体の意味特徴は「animate」である.

④ ϕ_3 : わたし

ϕ_3	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
外		×			
げんかんの戸		×			
かわいい子ねこ				×		..
わたし ▽				○		.

SELE: 述語「入れてやる」の主体 (が格) の意味特徴は「animate」である.

④ ϕ_4 : 子ねこ

ϕ_4	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
外		×			
げんかんの戸		×			
かわいい子ねこ ▽			○	×		..
わたし			×	○		.

SELE: 述語「入れてやる」の対象 (を格) の意味特徴は「animate」である.

DISJ: 述語「入れてやる」の主体 (が格) と対象 (を格) は異なる.

② ϕ_1 では, 制約CTOP (CTOPIC) の違反が, ④ ϕ_4 では, 制約DISJの違反とCTOPの違反の比較評価が働いている.

ここまでで設定された、ゼロ代名詞の解決に関わる制約を示すと次のようになる。DIFFERとCTOPICが新しく加えられている。SUBJECTを拡張したPROPSHを考えることができる。

DIFFER：述語の主体と文の発話者が異なる。

SELECT：先行詞の意味特徴はゼロ代名詞に対応する格フレーム要素の意味特徴と一致する。

DISJOINT：述語の共出現項は排他的である。

CTOPIC：副節の「が格ゼロ代名詞」は主題が設定されていれば主題を照応する。

SUBJECT：主格ゼロ代名詞は前文の主格先行詞を選好する。

[PROPSH：PROPERTY SHARING：先行詞と文法的な格属性を共有する.]

(SUBJECTの拡張，OBJECTの共有も含めて考える.)

DISTANCE：ゼロ代名詞と先行詞の距離はより短いほうがよい。

4. 制約の適用によるゼロ代名詞解析のモデル

3章で述べた最適性理論に基づくゼロ代名詞解析では、指示対象の候補集合の生成において、ゼロ代名詞が出現したところまでの文脈中のすべての対象を考えることになっている。これを制限するしくみと制約の適用を組み込んだゼロ代名詞解析のモデルについて考察を進める。

4.1 指示対象リスト

文脈や状況から生成され、ゼロ代名詞の指示対象の候補となる対象を保持するしくみとして指示対象リストを構成する。指示対象リストRL：Referent Listは、様々な構造のものが考えられる。最も素朴で簡単なものとして、CENTERモデルにおけるCENTER 1個から構成されるもの考えることができる^{1,2)}。本節では、CENTERモデルの拡張として、TOPIC（主題）を特別に取り立ててリストの要素とするモデルについても考える。そのようなモデルをCENTER++モデルと呼ぶことにする。また、成山によるSRL（Salient Referent List）は、より広い範囲から対象を選択してくるリストである^{4,5)}。それを改良した石田・藤井のモデル⁶⁾がある。本報告では、SRLと石田・藤井による改良を参照して、ORL（Ordered Referent List）を構成し、RLをTOPICとORLから構成されるものとするORLモデルについても考察する。それぞれの指示対象リストを次のように表記する。

・RL_C [CENTER]

・RL_T [TOPIC, cENTER]

・RL_O [TOPIC, ORL]

ゼロ代名詞が同定されたとき、関数Genで生成される候補集合をRLに制限することになる。

RLは順序構造も設定されていて、実は、これも一つの制約と考えることができる。中心性に基づく制約ということが出来る。制約チェックテーブルの最初の列に指示対象候補を並べるときにその優先順位が組み込まれることになる。

4.2 制約

ゼロ代名詞解析のために利用される基本的な制約については、すでに3.2でみたように Bouma et al. ¹³⁾は次のように分類している。

- ①構文的な情報 (syntactic information)
- ②意味的な情報 (semantic information)
- ③語用論的な情報 (pragmatic information)
 - salience (指示対象の活性度あるいは顕著性)
- ④世界知識 (world knowledge)

ここで、語用論的な情報の中に、salience (顕著性：中心性)が含まれている。これを制約とすると顕著性のランクをその値として区別する必要がある。本考察においては、これは指示対象リストの構成の中に組み込まれており、前節で述べたように、制約チェックテーブルの左端の列に指示対象の候補を並べるときに、そのランクの順に並べることにより組み込む。

DIFFER：述語の主体と文の発話者が異なる。

SELECT：先行詞の意味特徴はゼロ代名詞に対応する格フレーム要素の意味特徴と一致する。

DISJOINT：述語の共出現項は排他的である。

CTOPIC：副節の「が格ゼロ代名詞」は主題が設定されていれば主題を照応する。

SUBJECT：主格ゼロ代名詞は前文の主格先行詞を選好する。

[PROPSH：PROPERTY SHARING：先行詞と文法的な格属性を共有する.]

(SUBJECTの拡張, OBJECTの共有も含めて考える.)

DISTANCE：ゼロ代名詞と先行詞の距離はより短いほうがよい。

優先順位としては次のように考える。この妥当性についても再考する必要がある。

DIFF (ER) > SELE (CT) > DISJ (OINT) > CTOP (IC) > SUBJ (ECT) > DIST (ANCE)

4.3 ゼロ代名詞解析の基本的なモデル

4.3.1 CENTERモデル

文脈中の指示対象の保持のしくみを考えるとき、亀山によるCENTERモデル²⁾は、指示対象候補として一個の対象 (CENTER) を保持するモデルである。

【亀山によるCENTERモデル】

CENTER：発話における最も中心的な指示対象

[A] 発話文は、以下の順序（主題性ハイエラキー）で指示対象の中心性を決める。

[話題格（～は）>主格（～が）>目的格（～に／～を）>その他]

[B] 話題格「～は」は新しいCENTERを立てる。

[C] ゼロ代名詞は、その時のCENTERを指し、その中心性を続行させる。

この亀山によるCENTERモデルをゼロ代名詞の処理を行う基本的な単位である発話文を節として、また、中心性ハイエラキーの目的格を分けて、次のようなモデルとして構成し直しこれを本報告におけるCENTERモデル³⁾とする。

【CENTERモデル】

○文を左から右へ読み進む。

①話題格「～は」に出会ったら、その対象を新しいCENTERとする。

②節を単位として、ゼロ代名詞の同定を行う。

③ゼロ代名詞の照応要素をCENTERとする。

④ゼロ代名詞の照応要素を決定し、節の処理を終わったところで、その節で言及されている対象のリストを次の中心性ハイエラキー [H] に従って構成する。

[H][話題格（～は）>主格（～が）>目的格（～に）>直接目的格（～を）>その他]

この節対象リストLOC (List of Objects in Clause) のトップ要素をCENTERとして、次の節の処理へと進む。

複文の場合、含まれている節について、①から④の処理を繰り返し行う。

⑤文の処理を終わったところで、主節で言及されている対象のリストを中心性ハイエラキー [H] に従って構成する。この主節の対象リストのトップ要素を文のCENTERとして、次の文の処理へと進む。

【CENTERの決定方法】

○CENTERの初期設定：素朴なモデルとしては空である。

①CENTERは1個の対象である。

②話題格（～は）を特別に処理し、無条件にCENTERとする。

③節単位で節対象リストを構成する。

④節対象リストはゼロ代名詞の指示対象を解決したあとの節の構成要素から構成する。

その順序は次の中心性ハイエラキーによる。

[H]:[話題格（～は）>主格（～が）>間接目的格（～に）／直接目的格（～を）>その他]

⑤節対象リストの先頭（最左）の対象をCENTERとする。

【ゼロ代名詞の同定方法】

①述語（動詞、形容詞、判定詞）の格フレームの必須構成要素が表現されていない。

述語を読んで、述語の意味構造を構成するところでゼロ代名詞を同定する。

②名詞の意味フレームの必須構成要素が表現されていない。

名詞を読んで、名詞の意味構造を構成するところでゼロ代名詞を同定する。

【ゼロ代名詞の解決方法】

◇CENTERをゼロ代名詞の指示対象とする。

指示対象候補のリストはCENTER 1個から構成される。

RL_C : (CENTER)

【ゼロ代名詞解決結果の検証方法】

◇制約を適用することにより検証する。

具体例について検討する。

< 2 > ①外へ出ようと思って、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。

②わたしは、[ϕ_1 が]びっくりしましたが、

③ [ϕ_2 が]寒そうにしているので、

④ [ϕ_3 が][ϕ_4 を]家の中に入れてやりました。

①外へ出ようと思って、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。

RL_C : [かわいい子ねこ]

②わたしは、

RL_C : [わたし]

[ϕ_1 が]びっくりしましたが、

ϕ_1	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
わたし						.

RL_C : [わたし]

③ [ϕ_2 が]寒そうにしているので、

ϕ_2	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
わたし	×					.

制約DIFFの適用においては、「わたし」が「発話者」であることの判定が必要となる。

RL_C : [わたし]

④ [ϕ_3 が][ϕ_4 を]家の中に入れてやりました。

ϕ_3	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
わたし						.

ϕ_4	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
わたし			×			.

RL_C : [わたし]

RL_CはCENTER 1個なので、すべてのゼロ代名詞にそれが割り当てられることになり、その妥当性を制約によって検証することになる。上の例では、③ ϕ_2 と④ ϕ_4 がそれぞれ制約DIFF, DISJに違反している。これは妥当な解ではないという判定になる。CENTERモデルによる解析結果が妥当な解を与えない場合を検討することによって、ゼロ代名詞の解析に関わる構文的、意味的、語用論的な様々な制約を抽出することができる³⁾。

4.3.2 CENTER++モデル

指示対象の保持のしくみを段階的に拡張する第1歩として、CENTERモデルを拡張したCENTER++モデルを構成する。CENTER++モデルは、トピック（主題）を特別に取り立てるモデルである。

【CENTER++モデル】

①話題格（～は）を個別の指示対象候補とする。

それをTOPICとする。TOPICは、話題格（～は）の出現によって更新する。

②中心性のハイエラキーは、次のように設定する。

[H'] : [主格（～が） > 間接目的格（～に） > 直接目的格（～を） > その他]

③指示対象候補のリストは次のような構成とする。

RL_T : (TOPIC, cENTER)

CENTERと区別するために、cENTERと表記する。

cENTERは、[H']に従って、最も中心性の高いものに決定する。

④ゼロ代名詞の指示対象の決定は、RL_Tを指示対象候補リストとして、最左の要素TOPICに最も高い中心性を与えながら、最適性理論の考えに基づき、制約による評価により行う。

具体例について検討する。

< 2 > ①外へ出ようと思って、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。

②わたしは、[ϕ_1 が]びっくりしましたが、

③ [ϕ_2 が]寒そうにしているので、

④ [ϕ_3 が][ϕ_4 を]家の中に入れてやりました。

①外へ出ようと思って、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。

RL_T : [, かわいい子ねこ]

②わたしは、

RL_T: [わたし, かわいい子ねこ]

[ϕ_1 が] びっくりしましたが,

ϕ_1	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
T: わたし▽						.
かわいい子ねこ						..

RL_T: [わたし, かわいい子ねこ]

節の主格が「わたし」となり, TOPICと同じになる. このような場合は, cENTERを決定する節対象リストに加えないこととする. 「T:」はTOPICを表している.

③ [ϕ_2 が] 寒そうにしているので,

ϕ_2	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
T: わたし	×					.
かわいい子ねこ▽						..

制約DIFFの適用においては, 「わたし」が「発話者」であることの判定が必要となる.

RL_T: [わたし, かわいい子ねこ]

④ [ϕ_3 が][ϕ_4 を] 家の中に入れてやりました.

ϕ_3	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
T: わたし▽						..
かわいい子ねこ						.

ϕ_4	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
T: わたし			×			..
かわいい子ねこ▽						.

RL_T: [わたし, かわいい子ねこ]

RL_TはTOPICとcENTERの2個になる. 上の例では, 2個の情報を利用することによって, すべてのゼロ代名詞の解決が正しく行われている. ② ϕ_1 と④ ϕ_3 では, 指示対象候補リストの優先順位の高い対象を決定している.

4.3.3 ORLモデル

CENTER++モデルでは, 節中に現れる対象について, TOPIC以外は, cENTERだけを指示対象リストに保持している. 本節では, 節中に現れるすべての対象を指示対象として保持するリストとして, ORL (Ordered Referent List) について検討する. このリストは, 指示対象のリストとして, 成山によるSRL (Salient Referent List) アルゴリズム^{4,5)}, 石田・藤井による成山のアルゴリズムの改良版⁶⁾における指示対象のリストを参照して構成されている.

【ORLモデル】

- ①話題格（～は）を個別の指示対象候補とする。

それをTOPICとする。TOPICは、話題格（～は）の出現によって更新する。

- ②中心性のハイエラキーは、次のように設定する。

[H]:[主格（～が）>間接目的格（～に）>直接目的格（～を）>その他]

- ③ORL（Ordered Referent List）を次のように構成する。

節中に現れる指示対象を[H]に従って並べる。

さらに次のような動的な仕組みを組み込む。

[a] ゼロ代名詞の解決が終わったあとのゼロ代名詞の照応要素である指示対象もORLに加える。

[b] ORLの対象は同じ文法役割を持つ対象が出てくるまで継続して保持する。

- ④指示対象候補のリストは次のような構成になる。

RL_O : (TOPIC, ORL)

- ⑤ゼロ代名詞の指示対象の決定は、RL_Oを指示対象候補リストとして、最左の要素TOPICに最も高い中心性を与えながら、最適性理論の考えに基づき、制約による評価により行う。

ORLは、節が処理されるごとに、その節に出現する対象をそれまでに構成されているORLにマージすることによって更新される。その対象がORLにすでに存在している対象と同じ文法的な役割を担っているときは、その対象ですでに存在している対象を置き換える。成山によるSRLモデルでは、文を読み終わって、そこに明示的に現れた対象からSRLを構成し、そのSRLを利用して、その文に含まれるゼロ代名詞の解決処理を行う。ORLモデルでは、節を読み終わって、その節の前の節から構成されているRL_Oを利用してゼロ代名詞の解決処理を行い、その後、節で言及された指示対象についてORLを構成し、次の節の処理へと進む。明示的な表現による対象だけでなく、節におけるゼロ代名詞の解決処理の結果である対象も節単位のORLの中に組み込んでいく。ポイントを整理すると次のようになる。

- ・ゼロ代名詞を含んでいる節に含まれる表現の指示対象は利用しない。
- ・一つ前の節とそれより前の節から構成される指示対象のリストを参照する。
- ・節の処理が終わった時点で、ORLを構成する。

その構成方法は文法的な役割に依存したマージである。また、ゼロ代名詞の解決の結果もマージする。ただし、その結果の指示対象が、TOPICと同じ場合は、マージしない。

- ・TOPICは出現した時点で保持する。

具体的な例について検討する。

<2>① [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。

- ②わたしは, [ϕ_3 が] びっくりしましたが,
 ③ [ϕ_4 が] 寒そうにしているので,
 ④ [ϕ_5 が][ϕ_6 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_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.

RL_O : [(TOPIC), かわいい子ねこ (が : animate), げんかんの戸 (を : inanimate)]

ϕ_1 , ϕ_2 の解決については, この文が第1番目の文であるということによる特別な処理を考へる必要がある. 成山⁹⁾, 石田・藤井⁹⁾では, 指示対象リストを利用した復元の前に, 次のような処理が設定されている.

- ・第1番目の文であって, リストに「～は」表示の指示対象が入っていない場合は, 一人称か総称をあてる.

ここでは, そのような方略は設定しないで処理を進めることにする.

また, 「外 (へ)」は説明の簡略化のために処理していない.

- ②わたしは, [ϕ_3 が] びっくりしましたが,

RL_O : [わたし (TOPIC : first person ; animate), かわいい子ねこ (が : animate),
 げんかんの戸 (を : inanimate)]

ϕ_3	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
T : わたし▽						.
かわいい子ねこ						..
げんかんの戸		×				...

RL_O : [わたし (TOPIC : first person ; animate), かわいい子ねこ (が : animate),
 げんかんの戸 (を : inanimate)]

[ϕ_3 が] は [わたしが] となり, TOPICと同じ指示対象である. これはマージしない.[かわいい子ねこ] をそのまま保持する.

- ③ [ϕ_4 が] 寒そうにしているので,

ϕ_4	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
T : わたし	×					.
かわいい子ねこ▽						..
げんかんの戸		×				...

RL_O : [わたし (TOPIC : first person ; animate), かわいい子ねこ (が : animate),
 げんかんの戸 (を : inanimate)]

④ [ϕ_5 が^s][ϕ_6 を] 家の中に入れてやりました。

ϕ_5	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
T:わたし▽						.
かわいい子ねこ						..
げんかんの戸		×				...

ϕ_6	DIFF	SELE	DISJ	CTOP	SUBJ	DIST
T:わたし			×			.
かわいい子ねこ▽						..
げんかんの戸		×				...

RL_O: [わたし (TOPIC: first person; animate), かわいい子ねこ (が: animate),
げんかんの戸 (を: inanimate), 家の中 (<場所>: inanimate)]

RL_OはTOPICとORLから構成される。上の例では、ORLの情報を利用することによって、すべてのゼロ代名詞の解決が正しく行われた。② ϕ_3 と④ ϕ_5 では、CENTER++モデルと同じように、指示対象候補リストの優先順位の高い対象から決定している。

5. おわりに

さらにより多くの日本語文章におけるゼロ代名詞の出現例について検討を進め、指示対象候補の保持のしくみと指示対象の妥当性を検証する、あるいは最適性を評価するための制約とその適用のしくみについて詳細な考察を進める必要がある。本考察は、ゼロ代名詞解析のための知識と文脈情報の利用のしくみを制約の適用というしくみとして実現することの可能性を探るための一つの基礎的な考察と考える。

謝辞

本研究の一部は、北海学園大学ハイテク・リサーチ・センター研究費による援助を受けて行われました。また、学部4年生松本貴博君にはデータの収集と解析に協力を得ました。

参考文献

- 1) 桃内佳雄: 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, 北海学園大学工学部研究報告, 27, pp.303-327, 2000.
- 2) 亀山恵: 談話分析: 整合性と結束性, 「田窪行則他著, 談話と文脈, 岩波書店」, pp.93-121, 1999.
- 3) 桃内佳雄, 柴田更紗: CENTERモデルによる日本語ゼロ代名詞解析に関する基礎的考察, 北海学園大学大学院工学研究科紀要「工学研究」, No. 5, pp.85-92, 2005.
- 4) S. Nariyama: Multiple argument ellipsis resolution in Japanese, Proc. of MT Summit VIII, pp. 241-245, 2001.

- 5) S. Nariyama : Grammar for ellipsis resolution in Japanese, Proc. of TMI-2002 Conference, pp. 135-145, 2002.
- 6) 石田邦子, 藤井聖子 : 自然会話における名詞句省略とその復元, 言語処理学会第12回年次大会発表論文集, pp. 981-984, 2006.
- 7) D. I. Beaver : The optimization of discourse anaphora, Linguistics and Philosophy, 27 (1), pp. 3 -56, 2004.
- 8) D. Byron and W. Gregg-Harrison : Evaluating optimality theory for pronoun resolution algorithm specification, Proc. of the Discourse anaphora and reference resolution conference (DAAR2004), pp. 27-32, 2004.
- 9) M. Hong : Optimality Theory in Semantics and the Anaphora Resolution in Korean : An Adumbration, Language and Information, Vol. 6 , No. 2 , pp. 1 -22, 2002.
- 10) G. Bouma : Doing Dutch Pronouns Automatically in Optimality Theory, Proc. of the EACL 2003 Workshop on the Computational Treatment of Anaphora, pp. 15-22, 2003.
- 11) G. Bouma, P. Hendricks, H. de Hoop, I. Kramer, H. de Swart, and J. Zwarts : Conflicts in Interpretation, KNAW 05 (Colloquim on Cognitive Foundations of Interpretation), 2005.
- 12) A. Prince and P. Smolensky : Optimality Theory, Blackwel Publishing, 2004.
- 13) 北原真冬 : 最適性理論, 言語科学の百科事典 (編集委員長鈴木良治), 丸善, 2006.
- 14) 木下順次, 松村明, 柴田武他 : しょうがくこくご 1 上・下 / 小学国語 2 上・下 ~ 6 上・下, 教育出版 (株), 1983.