

## 決定木学習による代名詞「自分」の照応解析

長崎英紀 古宮嘉那子 柴原一友 但馬康宏 小谷善行  
東京農工大学 情報工学専攻  
〒184-0012 東京都小金井市中町 2-24-16

照応解析は同じ実体を指す異なる 2 つの表現の対応関係を解析するというタスクである。本研究は特定の代名詞「自分」の照応規則を追求・解明することを目的とした。

決定木学習によって分類器を生成した。素性は代名詞「自分」の特徴を踏まえ、人称の判別、先行詞同定の各処理に対処できるように設計した。先行詞候補・「自分」それぞれに関する情報として、表層格、品詞、概念などを、先行詞候補と「自分」の関係を表現する素性として、統語関係、談話関係、順序関係などを使用した。

「自分」照応解析実験を行った結果、解析対象を細分化する有効性を示せたことと、生成された決定木から、代名詞「自分」の照応規則の抽出に成功したことを報告する。

## Anaphora Resolution of Japanese Pronoun “Jibun” Using Decision Tree Learning

Hideki Nagasaki Kanako Komiya Kazutomo Shibahara  
Yasuhiro Tajima Yoshiyuki Kotani  
Tokyo University of Agriculture and Technology  
2-24-16 Nakamachi, Koganei Tokyo, 184-0012, Japan

Anaphora resolution is a task to identify two different expressions that indicate one same entity. In this paper, the authors present a method to analyze anaphora resolution of a particular pronoun “Jibun”. The authors created a classifier using decision tree. Considering characteristics of “Jibun”, features were designed corresponding to person distinction and antecedent identification. The authors used features regarding antecedent candidate and anaphor, respectively, such as surface case, part of speech, concept, etc. They also used features regarding the relationship between an antecedent and an anaphor, such as syntactic relation, discourse relation, order relation, etc. From the test results of the anaphora resolution, the authors succeeded in extracting the rules of anaphor resolution of “Jibun”.

### 1. はじめに

自然言語処理における照応解析は同じ実体を指す異なる 2 つの表現を解析するというタスクである。本研究はその 1 つである代名詞「自分」に絞り、照応問題の解析を深く追及することを試みた。機械学習による照応解析の研究は McCarthy ら [McCarthy, 1995], Soon ら [Soon, 2001], Yang ら [Yang, 2003] 等によって行われてきたが、いずれも汎用的な名詞句同士の指示関係を解明するというアプローチである。しかし、このアプローチでは名詞句の種類別に存在し得る照応規則の差異を発見しにくいと、ある種類の名詞句の解析は有効に行えても、他の名詞句の解析は失敗する可能性がある。筆者らは、ある特定の一つの代名詞である「自分」に着目して解析を行い、その成果を利用し

て他の代名詞や指示詞に応用範囲を広げていくというアプローチをとる。扱う対象を細分化することで、問題点をより明確にすることができ、研究成果につながれると考える。

### 2. 代名詞「自分」の特徴

筆者は「自分」の人称をいくつかに分類できると考えた。本研究において、それらを“話し手”、“聞き手”、“第三者”、“人一般”の 4 種類の人称として分類できるとした [長崎, 2007]。“人一般”は特定の人物ではなく状況に応じて「その行為をしている人」を指すような場合である。不特定の人物と照応するので、先行詞を同定する必要がない。例文 1 では、「知る人」が「自分」の指す人物と推測されるが、この文が発せられた時点でその人物は不定であるので、

“人一般”となる。

それぞれの人称によって先行詞の有無や先行詞との照応の仕方が異なると考えられる。人称が話し手または聞き手の場合、同じ会話文の中であれば先行詞は代名詞である可能性が高い(例文 2, 例文 3)。一方、第三者の場合、先行詞は必ずしも代名詞とは限らず、固有表現として文章のあらゆる場所に存在する可能性がある(例文 4)。

(例文 1) → “人一般”

相手を知り、自分を知れば必ず勝つ、という意味だ。

引用元: [asahi, URL] 2007年02月13日

(例文 2) → “話し手”

「僕は自分を安売りしてない。」

[BARKS, URL]2006年10月17日

(例文 3) → “聞き手”

「あなたは自分で自分にメールを送ったことがあるだろうか？」

[読売, URL]2007年01月05日

(例文 4) → “第三者”

チャールズが、放送で自分のノーハウを見せてくれた。

[innolife, URL]

### 3. 「自分」の照応解析モデル

#### 3.1 照応解析の3つの処理

代名詞「自分」は柔軟性が高い代名詞であるため、人称代名詞として機能する場合と再帰代名詞として機能する場合がある。再帰代名詞として機能している場合、「自分」と同じ文の中に現れる可能性が高い。一方、人称代名詞として機能している場合、先行詞は「自分」と同じ文だけでなく文章全体のどこにでも現れる可能性がある。

従って、「自分」の照応解析を人称の判別を含め、3つの処理に分けて素性を設計した。人称判別および同一文内の先行詞同定には、「自分」と同じ文内の情報を素性に利用した。周辺文、すなわち文章全体の先行詞同定には周辺文の情報を素性に利用した(図1, 表3参照)。

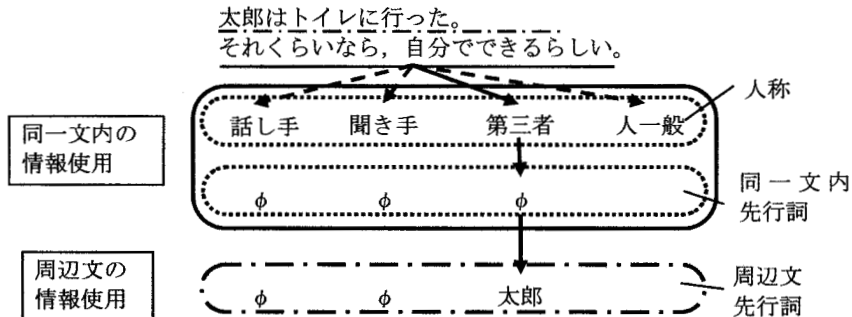


図 1: 照応解析の2つのフェーズ

#### 3.2 分類器の設計方法

単独候補モデル(Single Candidate Model

[Yang, 2003])で分類器を設計した。単独候補モデルは、ある先行詞候補に対して分類器が「自分」と照応するか否かを判断し、出力するモデルである。入力として解析対象の照応詞「自分」と先行詞候補に関する情報を与えると、分類器は出力として“照応”または“非照応”と解答する。

学習事例は正事例と負事例からなる。正事例は先行詞である名詞句、負事例は先行詞以外の名詞句である。下記のような例文があったとする。この文では「ナカ」が正事例である。「魔法」

や「所々」など、その他の先行詞候補は負事例である。

(例文 5)

ナカは自分の魔法を所々で見せてくれたね。

[スポーツナビ, URL] 2007年01月04日

また、この文から作られる事例は表 3-1 のようになる。

表 1: 単独候補モデルの学習事例

文 ID	名詞 (句)	・素性・	正解
39	ナカ	・素性値・	照応
39	魔法	・素性値・	非照応
39	所々	・素性値・	非照応

### 3.3 機械学習アルゴリズム：決定木学習

分類器を生成するための学習アルゴリズムには決定木学習アルゴリズム C4.5 を使用した。決定木を使用したのは、素性同士の関係を目で確認し、システムを改良することが可能だからである。

### 3.4 先行詞選定方法

● 先行詞候補の定義  
先行詞候補とは自分と照応する可能性のある名詞句の集合である。本稿において名詞句は名詞と同様の働きをする句であると定義する。また、名詞句は以下の条件を満たすものとする。

条件 1) 体言を含む文節において、同じ文節内に含まれる助詞以外の形態素全てを合わせて 1 つの複合名詞とする。

例) 元-メンバー (形態素 2 つ)

条件 2) 「A の B」の形の修飾節のみ 1 つの名詞句とする。なお、A と B は共に名詞または条件 1・2 を満たす名詞句である。

例) 半数近くの人

● 選定方法  
単独モデルでは出力値が“照応”か“非照応”の 2 値であるため、複数の候補が“照応”，すなわち先行詞と判定される場合がある。この問題を解決するために決定木の末端の確率を比較し、確率が最も高い候補を最終的な答えとして決定した。

### 3.5 照応詞と先行詞候補を抽出するための前処理

- テキストから文へ分割  
形態素解析および係り受け解析に入力するために、文章を文単位で区切った。1 文は地の文において句点 (。) または改行で区切られたものとした。
- 形態素解析  
形態素解析には茶筌[松本, 2003]を使用した。
- 係り受け解析  
係り受け解析には CaboCha[工藤, 2002]

を使用した。

- 名詞の結合  
前節において述べた先行詞候補の定義に従って名詞を結合し、名詞句を作成した。

- 先行詞候補のフィルタリング  
先行詞候補は文章のあらゆる場所に存在するが、真の先行詞は代名詞「自分」に近い位置など、存在する場所に偏りがあると推測した。「自分」を含んだ文章には約 100 個以上の先行詞候補が存在し、大量の負事例の中から正事例を選ばなければならない。局所部分以外の場所に存在する先行詞候補を除外することで、不要な負事例候補を削減することを試みた。この処理を先行詞候補のフィルタリングと呼ぶ。

フィルタリングの方法としては、「自分」が存在する文から前方 Bf 文から後方 Af 文以内の文に存在する候補だけ抽出する。

## 4. 素性の設計

実装した素性のうち、統語関係、ESRL、談話関係について述べる。

### 4.1 統語関係

先行詞候補と「自分」の文内での統語関係を以下のように定義し、学習させた。ここで用いる節の定義は、主語と述語動詞の組を持った要素であり、複文内に含まれる単文に相当する。  
同節内

先行詞候補が「自分」と同じ節の中にある主節  
「自分」の節が先行詞候補の節に属する(図 2)

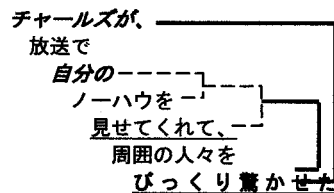


図 2: 主節の統語構造

### 従属節

先行詞候補の節が「自分」の節の一部である

### 連体修飾

先行詞の節が「自分」を連体修飾している

### 被連体修飾

先行詞候補が「自分」の節に連体修飾されている

### 並列節

先行詞候補の節と「自分」の節が互いに独立

して存在し、対等の関係にある

#### 4.2 ESRL

ある先行詞候補  $c_i$  が文章内で他の候補とどのような位置関係にあるかという情報も解析の手掛かりになると考えた。SRL[Nariyama, 2002]を拡張し、文章に含まれている全ての名詞句(すなわち先行詞候補)の格情報を保持する方法として Extracted-SRL (ESRL) を考案した。

ESRL は SRL と同様に5つのスロットを持つ。システムが文頭から文章を解析すると、対応する格のスロットに名詞句を順に格納していく。ただし、SRL と異なり、ESRL は各文ごとに5つのスロットを用意する。よって、ESRL のサイズは、「スロット5つ×文の数」となり、文頭から文末までスロットに格納された名詞句全てを格情報と共に保持する。

文章1があったとする。(a)~(c)の文のなかには5つの名詞句が含まれている。各名詞句はそれぞれの表層格と文番号に対応するスロットに格納される(図3参照)。

#### (文章 1)

	… …
(a)	…NP1 は NP2 を…
	… …
(b)	NP3 が, NP4 は…
	… …
(c)	…NP5 に……(φガ)V.
	… …

ESRL の情報から学習のために生成した素性について述べる。

#### 候補\_C 指標\_前

候補  $c_i$  と文章の先頭の間に、候補  $c_i$  よりもセンタリング指標[Nariyama, 2002]が高い他の候補の有無を示す素性である。図3の例において、点で塗りつぶされている部分に他の候補があれば素性値は「1」、無ければ素性値は「0」となる。

#### 候補\_C 指標\_後

候補  $c_i$  と「自分」を含む文の間に、候補  $c_i$  とセンタリング指標が同じ以上の他の候補の有無を示す素性である。図3の例において、斜線で塗りつぶされている部分に他の候補があれば素性値は「1」、無ければ素性値は「0」となる。

		文章の先頭	(a)	(b)	「自分」を含む文	(c)	文章の末端	
高 ↑ センタリング指標の順位 ↓ 低	係助詞(“ハ”)	..	NP1	..	NP4	..	NULL	..
	ガ格	..	NULL	..	NP3	..	NULL	..
	二格	..	NULL	..	NULL	..	NP5	..
	ヲ格	..	NP2	..	NULL	..	NULL	..
	その他	..	NULL	..	NULL	..	NULL	..
			c <sub>i</sub> =NP2の場合					

図3: ESRL (C 指標\_前および C 指標\_後で参照されるスロット)

#### 4.3 談話関係

扱うドキュメントによっては会話文が文の途中に挿入されている場合がある。本研究で使用したドキュメントはインターネット上のニュース記事であるが、例文6のような例が多数あった。この場合、自分と照応する名詞句は会話を発している江口さんである。会話文内にある自分の位置を基準にすると江口さんは外界の文脈に存在する先行詞であるが、地の文にある江口さんを基準にすると自分は同じ文内に存在

する照応詞である。すなわち、江口さんと自分の位置関係を同一文にあると捉えるか、異なる文にあると捉えるかという問題になる。

#### (例文 6)

江口さんは「開聞岳や錦江湾が一望できるきれいな所ですね。まさか自分が20万人目とは、うれしさでいっぱいです」と笑顔を見せていた。

【読売, URL], 2007年11月8日

そこで、先行詞と照応詞における文の談話構造上の関係を定義することで、この問題に対応することを試みた。まず 1 文の定義であるが、文章の地の文で句点（。）で区切ったものを大文、かぎ括弧（「」）で括られた会話文を句点で区切ったものを小文と定義した。大文と小文に先行詞と照応詞がそれぞれ属しているパターン毎に談話構造上の関係を表 2 のように定義した。談話関係は照応詞 ana に対する先行詞候補  $c_i$  の談話構造上の位置を表す。なお、この素性を候補\_照応詞\_談話関係と呼ぶ。

表 2：談話関係の定義

パターン	文の談話構造	談話関係
$c_i$ : 大文, ana : 小文	... $c_i$ ... 「... ana...」...	親文
$c_i$ : 大文, ana : 大文	... $c_i$ ... ana... 「...」...	同等
$c_i$ : 小文, ana : 小文	... 「 $c_i$ ... ana...」...	同等
$c_i$ : 小文, ana : 大文	... 「... $c_i$ ...」... ana...	子文

表 3：使用した全素性

	素性名	説明	素性の目的
先行詞候補に関する素性	候補_助詞	先行詞候補を含む文節の最初の助詞	同一文先行詞同定
	候補_品詞	候補の品詞（助詞の直前の形態素の品詞）	同一文先行詞同定
	候補_概念	候補の概念（EDR 概念辞書[EDR, 1995]の第5階層）	同一文先行詞同定
	候補_固有表現	候補が固有表現であるか否か	同一文先行詞同定
	候補_語彙分類	候補の語彙が「人間活動の主体」か否か（分類語彙表[国語研, 1996]を使用）	同一文先行詞同定
照応詞に関する素性	照応詞_接尾辞	「自分」に付く接尾辞	同一文先行詞同定
	照応詞_助詞	「自分」に付く助詞の表記（第1形態素+第2形態素）	人称判別
	照応詞_節_主語_1_人称	「自分」の節の1つ目の主語の人称	人称判別
	照応詞_節_主語_1_品詞	「自分」の節の1つ目の主語の品詞	人称判別
	照応詞_節_主語_2_人称	「自分」の節の2つ目の主語の人称	人称判別
	照応詞_節_主語_2_品詞	「自分」の節の2つ目の主語の品詞	人称判別
	照応詞_文_主語_1_人称	「自分」の文の1つ目の主語の人称	人称判別
	照応詞_文_主語_1_品詞	「自分」の文の1つ目の主語の品詞	人称判別
	照応詞_文_主語_2_人称	「自分」の文の2つ目の主語の人称	人称判別
	照応詞_文_主語_2_品詞	「自分」の文の2つ目の主語の品詞	人称判別
	照応詞_文_種類	「自分」の文の種類（会話文または地の文）	人称判別
	照応詞_文_述語_品詞 (×5)	「自分」の文の述語の品詞（先頭4つの形態素+末端の形態素）	人称判別
	照応詞_文_述語_活用型 (×5)	「自分」の文の述語の活用型	人称判別
照応詞_文_述語_活用形 (×5)	「自分」の文の述語の活用形	人称判別	
先行詞候補と照応詞に関する素性	候補_照応詞_統語関係	照応詞と先行詞候補間の統語関係	同一文先行詞同定
	候補_照応詞_順序	候補が「自分」の前に現れるか、後に現れるか	同一文先行詞同定
	候補_C指標_前	候補 $c_i$ 以前の文に候補 $c_i$ よりもセンタリング指標が高い他の候補があるか	周辺文先行詞同定
	候補_C指標_後	候補 $c_i$ より後の文に候補 $c_i$ とセンタリング指標が同じ以上の他の候補があるか	周辺文先行詞同定
	候補_照応詞_談話関係	候補が照応詞と談話構造上どのような関係にあるか	周辺文先行詞同定
	候補_照応詞_位置関係	候補が照応詞と同一文にあるか否か	周辺文先行詞同定

## 5. 評価実験

「自分」を含む文章を解析した。人称判別実験同様、本実験の事例データは全てインターネット上のニュース記事から収集し、人手によって正解の先行詞を記したものである。

実験は5分割交差検定を行い、ドキュメントファイル単位で解析した。(1つの「自分」を解析するために1つのドキュメントのデータを入力。)

### 5.1 評価指標

実データの解析タスクは1つの代名詞「自分」に対して1つの結果を出力することである。文章内に先行詞が何度も登場する場合は、解析対象である「自分」と照応するいずれかの先行詞を出力できれば正解とした。

「自分」が先行詞を持つ場合を正事例、先行詞を持たない場合を負事例とした。先行詞を持つ場合は「自分」=照応詞に相当し、先行詞を持たない場合は「自分」=非照応詞に相当する。照応解析の再現率、適合率、F値、精度はそれぞれ下記のように定義した。

$$\text{再現率} = \frac{\text{正しく先行詞同定した数}}{\text{照応詞の総数}}$$

$$\text{適合率} = \frac{\text{正しく先行詞同定した数}}{\text{"照応詞"と出力した数}}$$

$$F = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$$

$$= \frac{1}{2} (\text{"照応詞"と出力した数} + \text{照応詞の総数})$$

精度

$$= \frac{\text{正解先行詞同定数} + \text{正解の"非照応詞"回答数}}{\text{照応詞の総数} + \text{非照応詞の総数}}$$

また、「自分」の指示性、すなわち「自分」=照応詞か非照応詞かを区別できているかを計るため、指示性判別の再現率・適合率を以下のように定義した。

$$\text{再現率} = \text{指示性} S \text{の正答数} / S \text{の総数}$$

$$\text{適合率} = \text{指示性} S \text{の正答数} / S \text{の出力数}$$

(S: 照応詞 または 非照応詞)

実データ

インターネット上のニュース記事の379個のドキュメントから、775件の代名詞「自分」について解析した。先行詞候補の総数は Bf=4, Af=0 でフィルタリングを行ったとき 17,341 個であった。

ベースライン

ベースラインとして、「自分」の直前に出現する候補を先行詞として同定する簡易なルールベースシステムを作成した。このシステムでは、同一文内に「自分」より前方に候補が無い場合、1つ前の文にある最後の候補を選ぶ。

### 5.2 結果

先行詞候補のフィルタリングの最適な閾値を設定してから、照応解析実験を行った。Bf=4, Af=0 のときの照応解析結果を表4に、指示性判別結果を表5示す。また、情報利得比=0.02として決定木の枝刈りを行ったときの結果も示す。

表4: 「自分」照応解析実験結果

システム	再現率	適合率	F値	精度
本システム_枝刈り無し	24.3% (132/544)	35.9% (132/368)	28.9%	35.8% (271/758)
本システム_枝刈り有り	25.0% (136/544)	52.1% (136/261)	33.8%	40.1% (304/758)
ベースライン (直前の候補を出力)	15.1% (81/537)	10.9% (81/746)	12.6%	10.8% (81/747)

表5: 指示性判定結果

		再現率	適合率	F値	F値平均
本システム_枝刈り有り	照応詞 (先行詞あり)	39.5% (215/544)	82.4% (215/261)	53.4%	50.4%
	非照応詞 (先行詞なし)	78.5% (168/214)	33.8% (168/497)	47.30%	
ベースライン	照応詞	99.8%	71.8%	83.6%	41.8%

(直前の候補を出力)	(先行詞あり)	(536/537)	(536/746)	
	非照応詞 (先行詞なし)	0% (0/210)	0% (0/1)	0%

## 6. 考察

### 6.1 照応解析実験のベースラインおよび先行研究に対する成果

全ての実験においてベースラインを精度・F値とともに約10%以上上回ることができた。これは学習が適切に行われていることを示している。以下に飯田ら[飯田, 2005]の代名詞における照応解析の性能と本システムの性能を示す。飯田らのシステムとは解析対象および事例データが異なるため直接比較はできないが、ベースライン指標の一つとして考えることができる。飯田らのシステムは名詞句全般を解析対象としているのに対し、本研究は特定の代名詞に焦点を当てることで上回っている。名詞の種類別に照応規則を発見する重要性を裏付けることができた。

表 6: 先行研究との比較

	照応詞の種類	照応解析適合率
飯田ら	固有表現	84.3% (327/388)
	普通名詞	52.8% (254/481)
	代名詞	7.1% (1/14)
本システム _ 枝刈り有り	代名詞「自分」	52.1% (136/261)

### 6.2 生成された決定木について

生成された決定木の上位を図4に示す。この決定木は5分割交差検定で最も精度が高かったときの木である。候補の品詞、固有表現など、人であることを特定する素性と、談話関係・統語関係など文章の構造に関する素性が上位に現れている。「自分」の照応解析に重要な素性が判明した。

```

候補_品詞=名詞-固有名詞-人名
|<-yes—
| 照応詞_文_述語_活用型[1]=五段・マ行
| |<-yes—
| | “照応” 5件
| |—no->
| 候補_照応詞_談話関係=同等
| |<-yes—
| | “照応” 4件
| |—no->
| 照応詞_文_述語_品詞[3]=助詞-格助詞--

```

般

```

|<-yes—
| “照応” 4件
|—no->
| 照応詞_助詞=へ
|—no->
| 候補_品詞=名詞-接尾-人名
|<-yes—
| 照応詞_文_述語_活用形[4]=連用タ接続
| |<-yes—
| | “照応” 4件
| |—no->
| 候補_照応詞_統語関係=主節
| |<-yes—
| | “照応” 3件
| |—no->
| 候補_照応詞_順序=無し
|—no->
| 候補_固有表現=PERSON
|<-yes—
| 照応詞_文_主語_1_品詞=未知語
| |<-yes—
| | “照応” 5件
| |—no->
| 候補_助詞=無し
|—no->
| 候補_助詞=に対し
| |<-yes—
| | 候補_概念=無し
| |—no->
| 照応詞_文_主語_1_品詞=名詞-接尾-助数

```

詞

図 4: 生成された決定木 (枝刈り無し)

### 6.3 誤り例の分析

本システムは文章2の例で、正解先行詞である「**川本**」を分類器での判定時に「自分」と照応しないと出力している。このときに辿った決定木は図5のとおりである。候補が人名であるという素性のみが先行詞らしさの裏づけとなっているが、談話関係や照応詞の助詞など、この事例を解析するのに必要と推測される素性情報が木の中に現れていない。他の候補も全て「非照応」と判定されており、照応解析の結果として「自分」=「非照応詞」(先行詞なし)となっている。

談話関係については、同じ大文に含まれる候補同士は定義しているが、会話文から少し離れた地の文との談話関係を定義していないため、この例のような特徴を学習できなかった可能性がある。今後の課題としたい。

## (文章 2)

この試合、川本は一つもミスをしなかった。  
自分の守備の出来に満足だった。  
「負けは悔しいけれど、ミスしなかったのは  
3年間の練習の成果。自分で満足するプレー  
ができたのだから笑顔で終わりたい」

[asahi, URL]2007年07月12日

### 候補\_品詞=名詞-固有名詞-人名

```
|<-yes—
| 照応詞_文_主語 1_品詞=名詞-一般
| |—no->
| 候補_助詞=を, と, とともに
| |—no->
| 照応詞_文_述語_活用型[2]=特殊・マス
| |—no->
| 照応詞_節_主語 1_人称=無し
| |<-yes—
| | 照応詞_文_述語_活用型[末
| | 端]=特殊・タイ
| | |<-yes—
| | | “非照応” 2件
```

図 5: 先行詞誤答時に辿った木 (枝刈り無し) の経路

## 7. おわりに

文章内の全文から「自分」の先行詞同定と指示性判別を行った。名詞句の種類別の照応解析としては先行研究を上回った。

また、解析の具体例を検証することにより、他の自然言語処理の研究と深く関係する素性が分かった。談話関係を素性としてとり入れたが、本格的に導入するに当たっては談話構造解析の研究を熟知する必要がある。

照応解析の先行研究において未解決の問題であった、名詞の種類別に存在する照応規則の解明を代名詞「自分」について行うことができた。今後は本研究で解明した素性が更に他の再帰代名詞や人称代名詞の解析に応用され、自然言語処理の発展に寄与していくことを願う。

## 参考文献

[EDR, 1995]日本電子化辞書研究所: EDR 電子化辞書使用説明書, (1995).  
[McCarthy, 1995]McCarthy, Joseph F. and Lehnert, W. G.: Using decision trees for coreference resolution, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pp.1050-1055,

(1995).

[Nariyama, 2002]Nariyama, S.: Grammar for ellipsis resolution in Japanese, *Proceedings of the 9th International Conference on Theoretical and Methodological Issues in Machine Translation*, pp. 135-145, (2002).

[Soon, 2001]Wee Meng Soon, et al: A Machine Learning Approach to Coreference Resolution of Noun Phrases, *Computational Linguistics*, Vol. 27, No. 4, pp. 521-545, (2001).

[Yang, 2003]X. Yang, G. Zhou, J. Su, & C.L. Tan, Coreference resolution using competition learning approach, *Proc. 41st Annual Meeting of the Association for Computational Linguistics*, Vol. 1, pp. 176-183, (2003).

[飯田, 2005]飯田龍, 乾健太郎, 松本裕治, 関根聡: 最尤先行詞候補を用いた日本語名詞句同一指示解析, 情報処理学会論文誌, Vol. 46, No. 3, pp. 831-844, (2005).

[工藤, 2002]工藤拓, 松本裕治: Support Vector Machine を用いた Chunk 同定, 自然言語処理, Vol. 9, No. 5, pp. 3-21, (2002).

[国語研, 1996]『分類語彙表一増補改訂版一』付録CD-ROM, 国立国語研究所, (1996).

[長崎, 2007]長崎英紀, 古宮嘉那子, 但馬康宏, 小谷善行: 機械学習による代名詞「自分」の人称判別システム, 言語処理学会第13回年次大会 (NLP2007), pp. 59-62, (2007).

[松本, 2003]松本裕治, 北内啓, 平野善隆, 松田寛, 高岡一馬, 浅原正幸: 形態素解析システム『茶釜』version 2.3.3 使用説明書, 奈良先端科学技術大学院大学, (2003).

## 引用元コーパス

[asahi, URL] <http://www.asahi.com/>

[BARKS, URL] <http://www.barks.jp/news/>

[innolife, URL] <http://www.innolife.net/>

[スポーツナビ, URL]

<http://sportsnavi.yahoo.co.jp/>

[読売, URL] <http://www.yomiuri.co.jp/>