

## 日本語名詞における間接照応

村田 真樹 長尾 眞

京都大学大学院 工学研究科 電子通信工学専攻

## 1 はじめに

照応現象の一つに、文章中に現れていないがすでに言及されたことに関係する事物を間接的に指示する間接照応という用法がある[1]。たとえば、「家がある。屋根は白い。」の場合、「屋根」は前文の「家」の屋根である。間接照応の研究はこれまで自然言語処理においてあまり行なわれていなかったが<sup>1</sup>、文章の結束性の把握や意味理解において重要な問題である。そこで、我々は主題・焦点、表層表現、二つの名詞間の関係に関する知識などの情報を用いて日本語文章上でこの問題を解決する。間接照応の照応詞としては名詞句、指示詞、ゼロ代名詞が考えられるが、本研究は、名詞句が照応詞である場合の間接照応だけを対象とする。

## 2 間接照応の解析方法

間接照応の先行詞になりえる事物は、間接照応の照応詞によってある程度限定される。例えば、以下の例文のように「屋根」が照応詞である場合は、先行詞は「家」などの建物に限定される。

家がある。 屋根 は白い。 (1)

そこで、間接照応の解析を行なうには、間接照応の先行詞と照応詞の間の条件を記載した辞書が必要となる。

先行詞と照応詞の間の条件を記載するとき、先行詞でまとめて記載するか、照応詞でまとめて記載するかの問題がある。先行詞でまとめて記載すると関係の種類が爆発的に増加することになるが、照応詞でまとめて記載すると関係の種類数を少しに抑えることができる。例えば、「家族」が間接照応の照応詞である場合、先行詞としては「人」ぐらいであるが、「家族」が間接照応の先行詞である場合、照応詞としては「人

<sup>1</sup>文献[2]では化学の世界に限定して名詞「体積」の間接照応の解析をしているが、一般の名詞すべてに対して間接照応の解析を行なっている研究はない。

表 1: 名詞格フレーム辞書の例

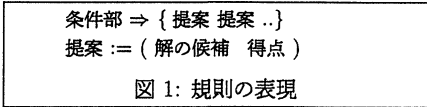
照応詞	先行詞となりえるもの	照応詞と先行詞の関係
家族	人	所属
国民	国	所属
元首	国	所属
屋根	建物	全体 - 部分
模型	生産物 (飛行機, 船)	対象
行事	組織	関与
人格	人	所有
教育	人 人 能力 (数学, 技術)	行為者 受益者 対象
研究	人, 組織 学問 (数学, 技術)	行為者 対象

数」「成員」「文化的水準」などと多様なものが想定される。よって、照応詞でまとめて記載の方が効率的であることがわかる。

照応詞でまとめると二つの名詞の間の関係に関する辞書の表現は動詞の格フレームに似たものになる。この名詞に対する辞書を名詞格フレーム辞書と呼ぶことにする。名詞格フレームの例は表1のようなものである。

ところが、このような辞書は今のところ存在しない。そこで、名詞格フレーム辞書の代わりに、照応詞が用言の派生語である場合は用言の格フレーム辞書を用い、照応詞が用言の派生語でない場合は、「名詞 A の名詞 B」の用例を用いる。

用言の格フレーム辞書を代用する場合はあまり副作用は生じないと考えられる。「名詞 A の名詞 B」を代用する場合はかなり副作用が生じると考えられる。そこで、以下の処理を施すことにより副作用を減らすことにした。(1) 名詞 A が「本当」などの形容詞的な名詞、数量表現、時間を示す表現である場合、その用例は用いない。(2) 名詞 B が「鶴」「人間」などのような間接照応の照応詞にならないものである場合、その用例は用いない。この二つ処理だけでは副作用を完全に消すことはできないが、少しは軽減できると期待で



きる。

以上の知識を用いて間接照応の解析を行なう。解析は以下の手順で行なう。(1) 解析する名詞に対して、「A の B」の用例と用言格フレーム辞書を用いて、間接照応先を求める必要のある空の要素を検出する。(2) 主題や焦点や主語から照応先の候補をあげる。(3) 「A の B」の用例と用言格フレーム辞書を用いて意味的制約から照応先を判定する。

以上の方法で一般の名詞における間接照応は解析できるが、「一部」などの部分を表わす名詞や「隣」などの空間語については特有の処理が必要となる。以下の例文の「一部」のように用言の格要素である場合は、その用言との意味的整合性の情報を利用する。

物資は水、戦車、弾薬が目につくが、おおいをかけられ積み荷が分からない車も多数ある。  
一部はさらに北西のラフハに向かい、積み荷を降ろしてダンマンに戻るトラックと行き交うたびに砂ばこりをあげていた。(2)

また、以下の例文の「隣」のように体言に係る場合は、その体言と意味的に近いものだけを先行詞とすることによって解析する。

お爺さんは大喜びをして家に帰りました。  
 そして、その夜起こったことを人々に話して聞かせるのでした。(3)  
 さて、隣の家に瘤のあるお爺さんがもう一人住んでおりました。

### 3 照応処理の枠組

#### 3.1 推定の手順

本研究では、名詞における間接照応の解析を行なう際、名詞、指示詞、代名詞、ゼロ代名詞などによる直接照応の解析も同時に行なう。本研究の照応解析は、照応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解の候補に得点を与えて、合計点が最も高い解の候補をシステムの解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 [3]。その結果に対して文頭から順に文節ごとにすべての規則を適用して照応解析を行なう。規則は、図 1 の構造をしている。

表 2: 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 3: 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	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

図中の「条件部」には文章中のあらゆる語やその分類語彙表 [4] の分類番号や IPAL の格フレーム [5] の情報や名詞の指示性の情報や構文解析・格解析の結果の情報などを条件として書くことができる。「解の候補」には先行詞となる名詞の位置を書くことができる。「得点」は解としての適切さの度合を表している。

#### 3.2 照応解析に用いる規則

名詞の解析のために規則を 15 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。

1. 解析している名詞の指示性<sup>2</sup>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 を直接照応の先行詞とする 30)}
2. 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)}
3. 名詞の指示性が特定性の不定名詞の場合 {(特定性の不定名詞として個体導入 10)}
4. 定名詞である度合が  $d$  で、用言からの派生語ではないが、間接照応の照応詞となる名詞 B の場合 (ただし、定名詞である度合  $d$  は文献 [6] で求めた指示性により適当に与える。)
 

{ (「名詞 A の名詞 B」の用例の名詞 A との類似度により与えられる得点が  $s$  で、重みが  $w$  で  $n$  個前にある主題を間接照応の先行詞とする  $w - n + d + s$ )  
 (用例との類似度により与えられる得点が  $s$  で、重みが  $w$  で  $n$  個前の焦点を間接照応の先行詞とする  $w - n + d + s$ )  
 (用例との類似度により与えられる得点が  $s$  で、解析している名詞に係る動詞の主格  $23 + d + s$ )  
 (用例との類似度により与えられる得点が  $s$  で、解析している名詞に係る動詞に係る名詞を間接照応の先行詞と

<sup>2</sup>名詞の指示性は文献 [6] の方法で推定する。

する  $23 + d + s$ }}

主題や焦点の定義と重み  $w$  は表 2, 表 3 のとおりである。「名詞 A の名詞 B」の用例の名詞 A との類似度  $s$  は、分類語彙表における名詞 A と先行詞の類似レベルに応じて与える。このとき、名詞 A が形容詞的な名詞である用例は利用しない。

5. 用言からの派生語の場合,  
 {(ゼロ代名詞解析モジュール [7] で解析する 20)}  
 ゼロ代名詞解析モジュールでは、主題、焦点、主語から候補をあげ、用言の格フレームの用例との類似度と、候補と照応詞の距離などから得点を与え、合計点が最も高い候補を先行詞とする。
6. 「一部」「隣」などの特殊な名詞で、助詞「の」がつく場合  
 {(今解析している名詞に係る名詞と同一の前方にある名詞を間接照応の先行詞とする 30)}

## 4 実験と検討

間接照応の解析を行なう前に構文解析・格解析を行なうが、そこで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。「名詞 A の名詞 B」の用例は EDR の共起辞書 [8] のものを利用した。間接照応の解析例を図 2 に示す。図 2 は名詞「食い違い」のガ格の解析を正しく行なったことを示している。同様に「食い違い」のト格の解析も行なわれる。本研究で提案した「名詞 A の名詞 B」の用例と用言格フレーム辞書を用いる方法で間接照応の解析を行なった結果を表 4 に示す。テストサンプルにおいても再現率 63%、適合率 68% の精度を得ているので、名詞格フレーム辞書が存在しない現在においてもある程度の精度で間接照応の解析ができることがわかる。

また、「名詞 A の名詞 B」と用言の格フレームを用いた近似的な方法による実験の他に、名詞格フレーム辞書を用いることができることを仮定した評価も行なった。この評価は、「名詞 A の名詞 B」と用言の格フレームを用いる近似的な方法で解析した結果において、以下の三つの理由で誤ったものを正解として数えることによって行なった。(1) 適切な用例が不足している。(2) 副作用を示す用例が存在している。(3) 名詞と動詞の場合で格フレームが異なる。実際に名詞格フレーム辞書を作って解析する場合は、辞書に誤りが含まれることが予想され、精度はここで示したものよりも若干低くなると思われる。

実験結果の中には名詞格フレーム辞書があったとし

太郎はデータをコンピュータに打ち込んでいました。  
 やっと半分やり終えました。  
 次郎が古いデータを見せました。  
 いくつかの食い違いを説明しました。

ゼロ代名詞解析モジュールでの得点状況

	古いデータ	次郎	埋めない
焦点	14	15	
意味素性		-5	
用例	4	-10	
埋めない			15
合計	18	0	15

「食い違い」の格フレーム

表層格	意味素性	用例
ガ	A B S (抽象名詞)	データ、意見、意識
ト*	A B S (抽象名詞)	データ、見方

図 2: 間接照応の解析例

ても本手法では扱えないものがあつた。

30日に竹下首相をまじえて、自民党の関係者が話し合った内容によると、1票の格差は3倍以内を目標に是正する、総定数を1減らして511に戻す、抵抗 (4) の強い2人区の解消は見送り、定数6の北海道1区だけ分区する方針で合意した、という。

この例文の下線部の「抵抗」の間接照応先は一見すれば「2人区」と思ってしまうが、深く考えれば「2人区の立候補者」であることがわかる。このような場合は、「2人区」から「抵抗」まで間接照応するとき、「2人区 — 立候補者」、「立候補者が抵抗する」の二段階の知識の利用が必要となる。このような場合も処理できるような枠組にしていく必要がある。しかし、知識の二段階の利用をあらゆるものに適用すると、間接照応でないのに誤って間接照応と解析する場合が増加し、適合率を低下させることになる。二段階の知識を利用する場合は、また新たな考え方が必要である。

## 5 名詞格フレーム辞書の作成に関する考察

名詞格フレーム辞書の作成コストについて考察する。「名詞 A の名詞 B」の意味解析の研究が進めば、名詞格フレーム辞書は自動的に作成することができ。この場合は非常に小さいコストで名詞格フレーム辞書を作成できることになる。「名詞 A の名詞 B」の意味解析の研究が今のような状態でなかなか困難で、人手で作成せざるをえない場合は、どうだろうか。このときも、「名詞 A の名詞 B」の用例に注目して作成するのがよいと考えられる。例えば、EDR の「名詞 A の名詞 B」の用例を名詞 B の分類語彙表の分類番号に応じて並べかえ、また、名詞 A の分類語彙表の分類

表 4: 本研究の実験結果

	用言からの派生名詞以外		用言からの派生名詞		合計	
	再現率	適合率	再現率	適合率	再現率	適合率
「名詞 A の名詞 B」と用言の格フレームを用いた実験						
学習サンプル	91%(60/66)	86%(60/70)	66%(23/35)	79%(23/29)	82%(83/101)	84%(83/99)
テストサンプル	63%(24/38)	83%(24/29)	63%(20/32)	56%(20/36)	63%(44/70)	68%(44/65)
名詞格フレーム辞書を用いることができる場合の評価						
学習サンプル	91%(60/66)	88%(60/68)	69%(24/35)	89%(24/27)	83%(84/101)	88%(84/95)
テストサンプル	79%(30/38)	86%(30/35)	63%(20/32)	77%(20/26)	71%(50/70)	82%(50/61)

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{ 例文 (43 文)[9], 童話「こぶとりじいさん」全文 (93 文)[10], 天声人語一日分 (26 文), 社説半日分 (26 文)}  
 テストサンプル{ 童話「つるのおんがえし」前から 91 文抜粋 [10], 天声人語二日分 (50 文), 社説半日分 (30 文)}  
 評価に適合率と再現率を用いたのは, 間接照応を行なわない名詞をシステムが誤って間接照応を行なうと解析することがあり, この誤りを適切に調べるためである。適合率は間接照応の先行詞を持つ名詞の要素のうち正解した要素の個数を, システムが間接照応の先行詞を持つと解析した要素の個数で割ったもので, 再現率は間接照応の先行詞を持つ要素のうち正解した要素の個数を, 間接照応の先行詞を持つ要素の個数で割ったものである。

表 5: 「名詞 A の名詞 B」を分類語彙表に基づいて並べかえたもの

名詞 B	名詞 A を分類語彙表に基づいて並べかえたもの
国民	(人間) 相手 (組織) 国 先進国 両国 内地 全国 日本 ソ連 英国 アメリカ スイス デンマーク 世界
元首	(人間) 来賓 (組織) 外国 各国 ボーランド
屋根	(組織) 北海道 世界 学校 工場 ガソリンスタンド スーパー 自宅 本部 (生産物) 車 住宅 家 邸宅 民家 神殿 玄関 車 車体 新車 (現象) 緑 オレンジ色 (動作) かかわらぶき (精神) 方式 (特徴) 形式 車体
模型	(動物) 象 (自然) 富士山 (生産物) 鋳物 マンション カプセル 電車 船 軍艦 飛行機 ジェット機 (動作) 造船 (精神) プラン (性質) 運行
行事	(人間) 皇室 王室 官民 家元 (組織) 全国 農村 県 日本 ソ連 寺 学校 学園 母校 (動作) 就任 まつり 祭り 祝い 巡礼 (精神) 祝い 恒例 公式
人格	(人間) わたし 私 人間 青少年 政治家

番号に応じて並べかえ, さらに, 名詞 A が形容詞的な名詞である用例を省くと表 5 のようになる。表 5 の形になれば, そこから名詞格フレーム辞書を人手で作成するのはそんなに大変なことではないと考えられる。

## 6 おわりに

本研究では, 名詞における間接照応の解析方法の提案を行なった。このとき, 名詞格フレーム辞書を作成することが望ましいが, 名詞格フレーム辞書はまだ存在しないので, 「名詞 A の名詞 B」の用例と用言格フレーム辞書を代わりに利用することにした。この方法で, テストサンプルにおいて再現率 63%, 適合率 68% の精度で解析できた。このことは, 名詞格フレーム辞書が存在しない現在においてもある程度の精度で間接

照応の解析ができることを意味している。また, 名詞格フレーム辞書が利用できることを仮定した実験も行なったが, この精度はテストサンプルにおいて再現率 71%, 適合率 82% であった。また, 名詞格フレーム辞書の作成に「名詞 A の名詞 B」を利用する方法を示した。

## 参考文献

- [1] 山梨正明, 推論と照応, (くろしお出版, 1992).
- [2] 長尾眞, 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] S. Kurohashi and M. Nagao, A method of case structure analysis for japanese sentences based on examples in case frame dictionary, *IEICE Transactions on Information and Systems*, Vol. E77-D, No. 2, (1994), pp. 227–239.
- [4] 国立国語研究所, 分類語彙表, (秀英出版, 1964).
- [5] 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).
- [6] Murata Masaki and Nagao Makoto, Determination of referential property and number of nouns in Japanese sentences for machine translation into English, *Proceedings of the 5th TMI*, (1993), pp. 218–225.
- [7] 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, 情報処理学会自然言語処理研究会 95-NL-108, (1995).
- [8] (株) 日本電子化辞書研究所, EDR 電子化辞書 日本語共起辞書第 1.0 版, (1995).
- [9] Marilyn Walker, Masayo Iida, Sharon Cote, Japanese Discourse and the Process of Centering, *Computational Linguistics*, Vol. 20, No. 2, (1994), pp. 193–232.
- [10] 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).