

日本語文章における名詞の指示対象の推定

村田 真樹 長尾 眞

京都大学工学部 電気工学第二学科

1 はじめに

日本語文章における名詞の指示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、本研究では主題・焦点、名詞の指示性、経験的規則などの情報を用いて名詞の指示対象を推定する。このとき指示詞や代名詞やゼロ代名詞の指示対象も推定する。日本語には冠詞がないことから、二つの名詞が照応関係にあるかどうかを判定することが困難である。これに対して、我々は冠詞に代わるものとして名詞の指示性[1]を研究しており、これを用いて名詞が照応するか否かを判定する。例えば、名詞の指示性が定名詞ならば既出の名詞と照応する可能性があるが、不定名詞ならば既出の名詞と照応しないと判定できる。さらに、名詞の修飾語や所有者の情報をを用い、より確実に指示対象の推定を行なう。

2 名詞・代名詞等の指示対象の推定方法

本研究では、以下で述べる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解析する。

表層表現から名詞の指示性を推定し¹これを利用して指示対象を推定する。指示性が定名詞句の場合は前方にある同一名詞を指示対象とする。指示性が定名詞句以外の場合は、前方の主題と焦点²から指示対象を探し、指示性の推定における定名詞句でない度合³と主

題・焦点の重みと指示対象との距離の三つの情報を組み合わせることににより照応するか否かを決定する。本来は定名詞句以外の場合は既出の名詞を指示することはないが、名詞の指示性の推定を誤ることがあり実際には定名詞句の可能性があるのでこのような処理を行なう。

修飾語を持つ名詞については、同じ修飾語を持つ同一名詞であることを照応する条件とする。例えば、以下の文章中の「左の頬」は修飾語の異なる「右の頬」と照応しない。

さて、隣の家に瘤のあるおじさんがもう一人住んでおりました。
このおじさんの瘤は右の頬にありました。(1)
(中略)
天狗達は、前の晩に来たおじさんから取った瘤をそのおじさんの左の頬に付けてしまいました。

所有者が推定できる名詞の場合は、同じ所有者を持つ同一名詞であることを照応する条件とする。例えば、以下の文章中の「頬」は所有者が同じ「おじさん」であることから照応する。

さて、おじさんには (おじさんの) 左の頬に瘤がありました。
それは人の拳ほどもある瘤でした。(2)
まるで (おじさんの) 頬を (おじさんが) 膨らませているかの様に見えるのであります。

所有者の推定は、意味素性⁴が動物の一部を意味するPARである名詞に対してのみ行なう。その名詞が存在する文の主語かそれまでの主題の中から意味素性がHUM(人間)かANI(動物)のものを探し出して、それを所有者とする。

以上は同一名詞の場合に限って説明したが、以下の例のようにある名詞を末尾に含む名詞がその名詞の指示対象となることがある。

オーストリアのレルヒ少佐が、日本の陸軍将校に、本格的にスキーを教えたのは、明治の末のことである。少佐は日本の軍人たちに前にしてまず「メテレスキー」と号令した。(3)

このような場合も解析できるように、本研究では以上の方法での名詞が同一であるという条件を末尾に含むという条件に変更する。ただし、末尾に含むという条

合が小さいほど照応しやすくなる。

⁴本研究では名詞意味素性辞書[3]を用いる。

¹ 本研究では名詞の指示性として総称名詞、定名詞、特定性不定名詞、不特定性不定名詞を考えている(総称名詞、定名詞、不定名詞の定義は文献[1]を参照のこと。特定性不定名詞句は話者が指示対象を認識している不定名詞句とし、不特定性不定名詞句は話者が指示対象を認識していない不定名詞句とする。[2])。総称名詞、定名詞、不定名詞の判定は文献[1]で行ない、不定名詞における特定性・不特定性の判定は“(名詞A)が存在しない”ならば名詞Aは不特定性”という表層表現を利用した決定的な規則により行なっている。文献[1]の利用においては以下のように規則を変更した。同一名詞が前方にある場合定名詞などに得点を与える規則があったがこれを省いた。また、意味素性[3]がPARの場合定名詞に得点を加えるという規則を追加した。

² 本研究での主題と焦点はそれぞれ表1、表2により定義する。

³ 文献[1]による指示性の推定では得点を用いており、定名詞句でない場合と推定した場合得点から定名詞句でない度合が得られる。その度

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

図 1: 候補列挙規則の表現

条件部 ⇒ (得点)

図 2: 候補判定規則の表現

件の場合は同一名詞の場合に比べ照応しにくくなるようにしておく。

指示詞については種類が多く、それぞれに対して詳細に規則を作ることによって指示対象を推定する。代名詞は会話文章中によく現れるので、会話文章の話し手や聞き手を把握することで⁵、指示対象を推定する。ゼロ代名詞は、主題や焦点と格フレームによる選択制限によって推定する [4]。

3 名詞等の指示対象を推定する枠組

3.1 推定の手順

本研究での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 [5]。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指示対象を推定する。規則には、指示対象の候補をあげるための候補列挙規則とその列挙された複数の候補に対して適用する候補判定規則の二種類がある。候補列挙規則は図 1、候補判定規則は図 2 の構造をしている。

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 [6] の分類番号や IPAL の格フレーム [7] 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。「指示対象の候補」には指示対象の候補とする名詞の位置もしくは「特定指示として導入」などを書くことができる。「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。これは特定性不定名詞句¹など

⁵ 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二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。

のように、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している。

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。まずすべての候補列挙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は加えてまとめる。次に列挙された指示対象の各候補に対してすべての候補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個ある場合は、一番初めに出された指示対象の候補を指示対象とする。

3.2 指示対象の推定に用いる規則

名詞の解析のための規則

名詞の解析は候補列挙規則のみで行なった。候補列挙規則は 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。

名詞の解析のための候補列挙規則

1. 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 (ただし、固有名詞の場合は修飾語や所有者の条件を無視する。また、末尾に限らず含まれればよいとする。)
{ (名詞 A 20) }
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) }⁶
3. 名詞の指示性が不特定性の不定名詞の場合
{ (不特定指示として個体導入 10) }⁶
4. 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) }
5. 指示性が定名詞以外の場合に適用される。d は文献 [1] によって推定した指示性により決まる値である。定名詞の得点を越える得点を総称名詞と不定名詞が持たない時、 $d = 0$ 。定名詞の得点より 1 点高い得点を総称名詞か不定名詞が持つ時、 $d = -3$ 。定名詞の得点より 2 点高い得点を総称名詞か不定名詞が持つ時、 $d = -6$ 。定名詞の得点より 3 点以上高い得点を総称名詞か不定名詞が持つ場合はこの規則は適用されない。
{ (修飾語や所有者が同じで重みが w で n 個前⁷の同一名詞の主題 $w - n + d + 4$)
(修飾語や所有者が同じで、今解析している名詞を末尾に含む重みが w で n 個前の主題 $w - n - 5 + d + 4$)
(修飾語や所有者が同じで重みが w で n 個前の同一名詞の焦点 $w - n + d + 4$)
(修飾語や所有者が同じで、今解析している名詞を末尾に含む重みが w で n 個前の焦点 $w - n - 5 + d + 4$) }
主題や焦点の定義と重みは表 1、表 2 のとおりである。

⁶ 総称指示もしくは不特定指示として導入された個体は、他の名詞から指示されないようにしている。

⁷ 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して数える。そうでない場合はそのままの位置で数える。

表 1: 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

指示詞や代名詞やゼロ代名詞の解析のための規則

指示詞や代名詞やゼロ代名詞を解析するため規則を100個ほど作成したが、そのうち主要なものを以下に示す。

指示詞や代名詞やゼロ代名詞の解析のための候補列挙規則

- 「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
{(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続補助詞か条件形を含む用言がある場合はその用言 15)}
- 名詞形態指示詞か「その / この / あの」の場合
{(重みが w で主題と焦点を合わせて数えて n 個前にある、同一文中か前文の主題 $w - n - 2$)
(重みが w で主題と焦点を合わせて数えて n 個前にある、同一文中か前文の焦点 $w - n + 4$)}
- 一人称の代名詞の場合 {(話し手 25)}
- ガ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 $w - n + 1$)}

指示詞や代名詞やゼロ代名詞の解析のための候補判定規則

- 「ここ / そこ / あそこ」であって、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。
- ソ系の連体詞形態指示詞の場合に、それが係る名詞 B の用例「名詞 A の名詞 B」⁸ 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。
- 代名詞の場合に、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。
- 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 [4]。

3.3 名詞の指示対象の推定例

名詞の指示対象を推定した例を図3に示す。これは図中の下線部の「火」の解析を正しく行なったことを示している。

4 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、そこでの誤りは人手で修正した。格フレーム

⁸この用例には EDR の共起辞書 [8] を用いる。

表 2: 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

その時お爺さんはあまり速くない所にある空き地に火が燃えているのに気が付きました。
赤い顔をして、鼻の青い、恐ろしい目付きの五六人の男が、火の周りに立っているのを見ました。

候補	(前文の)火	総称指示として導入
2番目の規則		10
5番目の規則	12	
合計	12	10

指示性の推定結果

指示性	不定名詞	定名詞	総称名詞
得点	1	2	3
5番目の規則	$= w - n + d + 4$ $= 15 - 4 - 3 + 4 = 12$		

図 3: 名詞の指示対象の推定例

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3に示す。名詞の解析精度は文中に指示対象が存在する名詞についてのものである。これは照応する名詞に注目したためである。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。

また、本稿であげた各手法の有効性を確かめるために指示性の利用の仕方をかえて表4の対照実験を行なった。表のように、本研究の規則による方法では適合率と再現率がともに均等に良かった。これは本研究の規則が指示性を適切に利用していることを意味している。指示性が定名詞句と推定された名詞句のみが照応するとした方法では再現率が悪い。これは指示性の推定の時に定名詞句であるのに他の名詞句と誤って推定し、照応しないとシステムが解析したためである。また、同一名詞はすべて照応するとした方法では、適合率が悪い。これは、同一名詞があってもそれらは照応するとは限らないということを意味している。

表 3: 本研究の実験結果

テキスト	文数	名詞の解析	指示詞の解析	代名詞の解析	ゼロ代名詞の解析
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 { 例文 (43 文), 童話「こぶとりじいさん」全文 (93 文) [9], 天声人語一日分 (26 文), 社説半日分 (26 文), サイエンス (16 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 [9], 天声人語二日分 (50 文), 社説半日分 (30 文), サイエンス (13 文) }

表 4: 名詞の解析における対照実験の結果

指示性が定名詞句と推定された名詞句のみ照応する	本研究の規則	指示性を用いない	修飾語・所有者の条件を用いない	修飾語・所有者の条件のみ用いる	同一名詞はすべて照応する	末尾に含む名詞とすべて照応する
学習サンプル						
92%(117/127)	82%(130/159)	72%(123/170)	65%(138/213)	64%(131/205)	52%(134/260)	47%(130/279)
76%(117/153)	85%(130/153)	80%(123/153)	90%(138/153)	86%(131/153)	88%(134/153)	85%(130/153)
テストサンプル						
92% (78/85)	79% (89/113)	69% (79/114)	58% (92/159)	58% (103/178)	47% (102/218)	44% (106/240)
68% (78/115)	77% (89/115)	69% (79/115)	80% (92/115)	90% (103/115)	89% (102/115)	92% (106/115)

表の上段と下段はそれぞれ適合率と再現率を表す。評価に適合率と再現率を用いたのは、先行詞がない名詞をシステムが誤って先行詞がある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。適合率は先行詞を持つ名詞のうち正解した名詞の個数を、システムが先行詞を持つと解析した名詞の個数で割ったもので、再現率は先行詞を持つ名詞のうち正解した名詞の個数を、先行詞を持つ名詞の個数で割ったものである。「同一名詞はすべて照応する」以外の推定では、先行詞は今解析している名詞を末尾に含むものであることを条件としている。「指示性が定名詞句と推定された名詞句のみ照応する」は名詞の解析のための 5 番目の規則をけずったものに相当し、「指示性を利用しない」は 1 番目の規則をけずり、5 番目の規則の d をけずりこの規則がいずれの指示性の時でも適用されるようにしたものに相当する。「修飾語・所有者の条件のみ用いる」は 1 番目の規則がいずれの指示性の時でも適用されるようにし、5 番目の規則をけずったものに相当する。

修飾語句や所有者を利用して指示対象の絞り込みを行なったが、これが有効に働いている。しかし、修飾語句が異なっても照応する場合があり、このような場合は解析を誤った。

そこでおじいさんは 近くの大きな杉の木の根元にある穴 で雨宿りをするにしました。
 (中略) (4)
 次の日、このおじいさんは山へ行って、杉の木の根元の穴を見つけました。

この例の下線部の「穴」は同一の穴であり照応するが、修飾語の文字列が異なっているため照応しないと誤って解析された。このような場合についても解析できるようにするには、異なる表現であっても同じ意味であることを把握できるようにする必要がある。

5 おわりに

本研究での手法は主に名詞の指示対象の推定に名詞の指示性や修飾語や所有者を用いることであった。実験を通じて、これらを用いることの有効性を示した。名詞の指示性の推定精度が向上すると名詞の指示対象の推定精度が向上すると考えている。そこで、名詞の指示性の推定精度を向上させる研究を行なう必要がある。

る。

参考文献

- [1] M. Murata and M. Nagao, 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.
- [2] 井上和子, 山田洋, 河野武, 成田一, 名詞, 現代の英文法, 第 6 巻, (研究社, 1985).
- [3] 渡辺靖彦, 黒橋慎夫, 長尾眞, 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).
- [4] 村田真樹, 長尾眞, 用例を用いた日本語文章におけるゼロ代名詞の指示対象の推定, 「IPAL シンポジウム'95」論文集, (情報処理振興事業協会 技術センター, 1995), pp. 63-66.
- [5] 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.
- [6] 国立国語研究所, 分類語彙表, (秀英出版, 1964).
- [7] 計算機用日本語基本動詞辞書 IPAL(Basic Verbs) 説明書, (情報処理振興事業協会技術センター, 1987).
- [8] (株) 日本電子化辞書研究所, EDR 電子化辞書 日本語共起辞書評価版第 2.1 版, (1994).
- [9] 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).