

日本語理解システムのための視点抽出と照応解決

清水一澄^{†,*} 横尾英俊[†]

文の連続体である談話では、省略や代名詞化などの照応現象が頻繁に利用される。そのような照応の具体的指示対象を明示的に示す照応理解あるいは照応解決と呼ばれる問題は、自然言語理解にとって一つの重要な課題である。本論文では、日本語談話の省略や代名詞化などの照応現象に対し、その指示対象を明示的に示す照応解決の一手法として、談話法上の制約を使った方法を提案する。提案する方法は、日本語文章の開始の2文を主に対象としている。本論文では、談話の開始の2文を助詞の「は」と「が」の使い分けによって、主語照応、新主題、反復主題、異主題という4種に分類する。これらの各種類に属する談話と各文の視点の関係を整理し、談話の第1文から次の文の主題になりやすい要素を予測する。本論文では、これを第1文発話直後の焦点と定義する。これを検出することによって、第2文で省略や代名詞化などの照応現象がある場合に、実際の照応の有無とは独立にそれを推定することができる。また、これに基づき実現した日本語照応理解システムについて紹介する。

Empathy Extraction and Anaphoric Disambiguation in Japanese Language Understanding System

KAZUKIYO SHIMIZU^{†,*} and HIDETOSHI YOKOO[†]

This paper presents a computational model for the resolution of anaphoric ambiguities in Japanese discourses such as stories. The proposed model deals mainly with the initial two sentences of each text. In the model, four types of initial two sentences are distinguished with respect to the use of the Japanese particles *wa* and *ga*. Those types are then used to select a plausible constituent of the first sentence which may serve as an anaphoric antecedent of the second sentence. This selection is based on the relation between the type and the *empathy*, which is the speaker's identification with a participant in the text. This paper also includes an actual implementation of the model using Common LISP.

1. はじめに

文の連続体である談話では、省略や代名詞化などの照応現象が頻繁に利用される。そのような照応の具体的指示対象を明示的に示す照応理解あるいは照応解決と呼ばれる問題は、自然言語理解にとって一つの重要な課題である。本論文では、主として次のような日本語談話に対して効果的な照応解決の手法を提案する。

- (1) 想定している談話とは、具体的な個々の事実に関する記述や物語のような文章である。
- (2) 談話の開始の2文を対象とする。
- (3) 第1文は、主題や主語を示す助詞の「は」、「が」を同時には含まないものとする。

本論文は、既に報告済みのモデル¹⁾に改良を加え、さらにプログラムとして実現を試みた結果の報告であ

る。

談話の種類を上述のように限定しても、特に日本語では、省略や代名詞化などの照応が多用される。そのような照応の解決には、文法的な知識から高度な常識等までを広範囲に利用した推論が必要になる。そのような中から、本論文では、視点や焦点などの談話法的な概念を利用した中間レベル^{**}の照応解決の手法を考察の対象にする。特に、提案する手法では、次のような方針を採用している。

- (4) 文法的知識（ある程度の意味的知識を含む）だけでは照応理解に複数の可能性が生じる場合を談話法的な制約によって解決する。
- (5) 代名詞などの照応詞（anaphor）の出現を見て初めて照応解決のメカニズムを動作させるのではな

[†] 群馬大学工学部情報工学科
Department of Computer Science, Gunma University

* 現在、シャープ株式会社
Presently with Sharp Corporation

** 名詞の性・数を利用するような低次のレベルと常識や世界の知識を利用する高次の推論レベル²⁾との中間に位置する照応解決法。このようなレベルによる照応理解は、より強い文脈によってくつがえされることは十分にありうる³⁾。

く、照応が実際に使用されるか否かにかかわらず、照応の対象となりやすい要素をあらかじめ明らかにする。

上の第2の方針(5)は、それを談話の開始の2文に適用した場合、第2文の照応を解決するとき既に第1文で照応の対象の候補を絞り込んでおくということに対応する。これは第1文で“focus”を仮定し、第2文で実際にそれが照応表現された場合に“focus”として確定するという Sidner⁴⁾ のモデルに相当している。実際、本論文でも、文の焦点をその文だけから決まる概念ではなく、次の文との関係で確定するものであるという立場をとっている。一方、同じ Sidner の流れを汲む日本語の照応解決のモデルの代表例としては、Kameyama⁵⁾ の“Centering”が有名である。ここでも、話し手の視点(Kameyamaは「自己同一視化(identification)」と呼んでいる)といった談話法的な概念が重要な役割を果たしている。しかし、視点の同定法自体についてはほとんど議論されておらず、Centeringが照応解決の枠組としては非常にすぐれたものであるにもかかわらず、結果として不十分なものとなっている。本論文で利用する視点などの概念は、主に久野の考察^{6)~8)}に基づいている。同じ久野に基づく類似のシステムとしては、山村ら⁹⁾の照応解決のモデルをあげることができる。しかし、本論文で示すような視点と焦点との区別が行われていないために、やはり、満足な結果が得られない例が種々観察される。

以上のような既存のモデルが必ずしも満足に働かない談話としては、上述の方針(5)を採用した場合でも第2文の微妙な差異が照応に影響するという、次のような例が典型的である。

(D 1) 1. 昨日、令子は街で貴子を見かけた。

2. 彼女は、あわてて追いかけた。

(D 2) 1. 昨日、令子は街で貴子を見かけた。

2. 彼女は、とてもきれいに着飾っていた。

この例では、(D 1-2)の「彼女」を「令子」と取るのが自然であるのに対し、(D 2-2)の「彼女」は「貴子」と考える方が自然である。本論文では、このような違いに関わる視点の役割と視点の認定法とを照応解決の枠組の中で議論する。提案するモデルは、既存のモデルの不十分な点を補強するものにはなっていないが、既存の枠組自体を否定するものではない。したがって、たとえば Centering の枠組の中で本論文のモデルを記述することも全く不可能というわけではな

い。しかし、本論文ではそのような立場をとらない。本論文では、このような立場を前提(1)~(3)との関係で議論する。

本論文の構成は次のとおりである。まず、次の2章で、本論文で中心的な役割を果たす視点と主題の認定法を議論する。主題は視点と同様、照応解決にとって重要な鍵となるものであり、Centeringの枠組でも既に利用されている¹⁰⁾。しかし、助詞の「は」によってマークされたものを主題として総括できるほど現象は単純ではない。ここでは、主題の使われ方を照応解決との関係で議論する。次の3章は、それまでの議論を総合した照応解決法の提案である。さらに、それが実現可能なモデルになっていることを4章の実システムの例によって示す。また、4章では、(1)~(3)を前提とした本論文の立場を考察する。

2. 焦点と視点

本章では、本論文で中心的な役割を果たす**焦点**、**視点**、**主題**という3概念について簡単に考察する。このような談話法上の概念は、厳密な定義を与えるのが必ずしも容易ではないものが多い。実際、各談話例において、それぞれの文の焦点等を一意に定めることが必ずできるわけではない。重要なのは、照応解決の手がかりの一つとしてこれらの現象を利用するのであって、それら自身の決定が最終的な目標ではないということである。

まず、**焦点**には、前節で述べた Sidner⁴⁾による特徴づけを含めて、次のような定義が考えられる。

(6) その文の要素の中で次の文の照応の対象に最もなりやすいもの。すなわち、次の文との関係で決まる。

(7) その文で初めて導入された新情報であり、前文から予測することの困難な要素。前の文との関係で決まる。

(8) その文で視点(後述)となっている要素の目に写っているその文の要素。

これらのうち、(7)が通常談話法等の言語学の方針で採用されている焦点の定義である。これに対し、本論文では、前節で述べた方針に従い、(6)の定義を採用する。それぞれの定義で決まる焦点は、実際の談話の上では一致する場合もあれば異なる場合もある。そのようなずれは、焦点の定義の困難さだけでなく、談話の時系列としての側面にも原因があると考えられる。本論文では、談話の開始の2文のうち「第1文の

焦点」を問題にする。これは、厳密には「第1文発話直後の焦点」と呼ぶべきものである。

以上の焦点の定義に従って、種々の談話の第1文を調べると、第1文の焦点は第2文の主題になる資格を有することが観察できる。つまり第2文で第1文の焦点を「～は」で受けることができる。例えば、次の(D3)では、第1文の「ほのかなやさしさ」を第2文で「それは」という主題にしている。主題は、この例の「それ」のように代名詞化することができる。また(D4)のように第2文で全く省略(ゼロ代名詞化)される場合もある。

(D3) 1. 冬服に着替えた日、ほのかなやさしさが私をつつんだ。

2. それは、しょうのうの匂いだった。

(D4) 1. ハイリブという猟師が白蛇を助けてやりました。

2. そして、宝の玉をお礼にもらいました。

このような例から、

(9) 第2文のゼロ化できる主題は第1文の焦点である。

という原則によって焦点を特徴づけることにする*。

次に視点について、この章ではこれまで述べた焦点とは独立に考察してみる。まず、久野⁷⁾にならって、文中の名詞句の指示対象(通常は人物)に対する話し手の自己同一視化の度合を共感度と定義する。もし、文中に話し手自身を指示する表現があれば、その共感度は最大値をとる。共感度は、この最大値以下の連続した値をとる量であるが、文中の最大共感度を与える要素 x の共感度が十分大きい場合に「話し手の視点は x にある」ということにする。したがって、ある文中のいずれの要素の共感度も十分大きくない場合には、その文には視点がないことになる。しかし、そのような場合でも、「視点がない」とは言わずに、「視点を明示する表現がない」ものと考え、次の原則を仮定する。

(10) 視点保存の原則

すでに仮定された視点がある場合、視点を明示する表現がない文の視点はその仮定された視点である。なお“視点を明示する表現がない文”とは、

- i) 以下で述べる「視点を明示する表現」を含まない文、および

- ii) 視点の推定規則で決まる視点の位置 (empathy locus^{6),10)} の格要素が照応化(ゼロ代名詞化など)されている場合、をいう。

この原則は、第1文に視点があって第2文に視点を明示する表現がない場合に適用する。

話し手が視点をおくのは、一般にその文の意味的な格要素となっているもので、動作主格 (Agent)、受益者格 (Beneficiary)、目標格 (Goal) などがある。「視点を明示する表現」としては、述部がその役割を果たす場合と名詞句が機能する場合とに大別できる。久野⁷⁾を参考に、そのような表現を整理すると次のようになる。以下の例では、2重下線が視点を表す。なお、「正の表現」とは、どこに視点があるかを積極的に示す表現となっていることを示している。これに対し、3章では視点がそこないことを示す「負の表現」を導入する。

(11) 述部による視点の推定 (正の表現)

- i) 述部に授受表現が存在するとき、

1. 述部に「(て)くれる」、「(て)もらう」があるなら、視点はその Beneficiary にある。

(例) 広は令子に宿題を手伝ってもらった。

2. 述部に「(て)やる」があるなら、視点は Agent にある。

(例) (D4)1. ハイリブという猟師が白蛇を助けてやりました。

- ii) 述部に方向表現が存在するとき、

1. 述部に「てくる」があるなら、視点は Goal (人物) にある。Goal が人物でなければ、視点はその文の記述内容の完了時点で Goal 側にいる人物にある。

(例) 突然、男がたかしになぐりかかってきた。

2. 述部に「ていく」があるなら、視点は Source にある。Source が人物でなければ Source 側にいる人物。

- iii) 述部が受動態なら、視点は主語 (人物) にある。

(例) りえはたかしに呼び止められた。

(12) 名詞句による視点の推定 (正の表現)

- i) 話し手を指す表現(「私」、「ぼく」など)が存在するなら、視点は話し手自身である。

- ii) 話し手自身を言及する表現が存在せず、固有名詞が存在する場合には、視点はその固有名詞の指す人物である。

ここまでの議論から明らかなように、視点とは談話

* この原則は、第1文の焦点の十分条件ではあっても、必要条件を述べたものではないことに注意。

の話手が自分と同一視しようとしている対象である。視点を問題にする以上、“話し手”自身の談話への関与の考慮は避けられない。しかも、話し手自身を指示する表現は、談話の第1文であっても省略されることが多いので更に検討を加える必要がある。次にこの点を、まず主題との関連で考察してみる。

主題は、すでに簡単に言及したように、「～は」を代表的な表現形式とし、「～」の部分について（すなわち、それを主題として）何かを述べるということを明示するものである。主題を提示する助詞「は」は、格助詞「が」との対比で論じられることが多い。一般に、「Xが」は新情報を表し、Xによらず省略することができない⁹⁾のに対し、「Xは」は省略可能な場合がある。そのような省略可能な場合として、まず、原則(9)に述べた第2文の「Xは」をあげることができる。一方、第1文に現れる「Xは」は次のようなものがある。

(13) 第1文の「Xは」の用法

- i) Xが話し手を指す表現である場合。
(例) ある日、ぼくは、小山に囲まれた不思議な場所を見つけた。
- ii) Xが「話し手+の+～」である場合。または、Xの埋め込みの構造の中に話し手自身を言及する表現がある場合。
(例) 私の父は大学に勤めています。
- iii) Xが固有名詞である場合。
(例) 広瀬さんたちは、奈良へ遊びに行きました。
- iv) Xが普通名詞であって、特定の個体（[+specific]）でなく、[-specific]あるいはタイプ¹¹⁾に分類されるものを指示する場合。
(例) 塩は私たちにとって一日も欠かすことができません。
- v) Xの対象が確定していて固有名詞に準ずると考えられる場合。
(例) 日本語は文字の種類が多い。

以上のほかにも、談話の種類を拡張すると「こちらは山田先生です。」のように「Xは」が何を指すかが発話場面の状況から確定できるときは、第1文で「Xは」を使うことがある。しかし、前提(1)という物語等の書き出しの文としては、このような例や普遍的な真実を述べる(13-iv)、(13-v)などの文が使われることはほとんどない。そこで以下では第1文に使われる主題文の種類として、(13-i, ii, iii)だけを考えることにする。これらは、いずれも第1文なので旧情報

を含むことがあり得ないにもかかわらず、「Xは」を使用することによって、それ以外の要素よりXを相対的に古い情報として位置付けようとする傾向が観察できる。これは常に可能なわけではなく、

1. Xが話し手自身（この場合「Xは」は省略されることが多い）であるか、
2. 話し手の視点がほとんどXに一致している場合などに限られる。このようなことから、「Xは」から始まる文、あるいは、その「Xは」が省略されたと考えられる第1文の視点については、既に述べた原則に加え次の仮定を行う。

(14) 「Xは」型第1文からの視点の推定

- i) 第1文に主題「Xは」があってXが固有名詞ならば、原則(11)、(12)にこの順で従う。
- ii) 第1文が主語「Xが」を有していないならば、視点は話し手にある。

この結果、

(D 5) たかしをほめてやった。

のような文が第1文にきた場合、原則(11)だけでは、(11-i-2)によって視点は不明となり、原則(12)では、「たかし」に視点があることになるのに対し、(14)を追加することで、正しく「話し手」に視点が設定される。

一方、「Xが」型の第1文では、上に述べたようにXが省略されることはないが、

(D 6) 先生がほめてくれた。

のようにガ格以外の要素が省略される場合がある。このような場合でも、第1文で省略可能なのは、一般に「話し手」だけであり、(14)に対応する規則として次を設定することにする。

(15) 「Xが」型第1文からの視点の推定

- i) 第1文に主語「Xが」があって、原則(11)で決まる視点の位置の格要素が空の場合には視点およびその格要素とも「話し手」である。
- ii) それ以外は、原則(11)、(12)にこの順で従う。

3. 視点を利用した焦点の決定

前提(2)に述べたように、本論文では、談話の開始の2文を特に考察の対象としている。第1文が全く文脈のない状態からの発話であるのに対し、第2文が第1文によって形成された文脈を前提としているという談話の典型的な構造が観察されるからである。例えば、主題を表す「は」の用法も第1文と第2文では前章で述べたようにある程度異なる。まずここでは、第1

文と第2文との関係を「は」、「が」の使い分けに着目して次の4種に分類する*。

- (a) **主語照応**
 (第1文) Xが
 (第2文) Xは
- (b) **新主題**
 (第1文) Xが……Y…….
 (第2文) Yは
- (c) **反復主題** (“は”のピリオド越え⁷⁾)
 (第1文) Xは
 (第2文) Xは
- (d) **異主題**
 (第1文) Xは……Y…….
 (第2文) Yは

以上の分類で、第2文の主題として書いた「Xは」や「Yは」は、実際の談話では、省略などの照応の対象となるものである。すなわち、これらは、第2文の主題であると同時に、第1文の発話直後の焦点でもある。特に、(b)と(d)は、第1文の主語や主題ではない要素が第2文の主題となる場合を示している。本論文の照応解決では、この場合のXとYとのいずれが第2文の実際の主題となっているかを文法的知識だけでは決められない場合を問題にする。例(D1)や(D2)の「令子」と「貴子」がそのような例にあたる。このような場合、YをXの対立候補と呼ぶ。一方、同じような例で「令子」と「たかし」が第1文に登場し、第2文の「彼女」に特別な曖昧さがないような場合には、文字どおり対立候補Yが存在しない場合と共に「Xに対する対立候補Yが存在しない」ということにする。

以上の4分類に属すと考えられる談話例からまず観察されることは、「(a)主語照応」と「(d)異主題」に相当する例の方が圧倒的に見つけやすいということである。これは、このような談話の出だしにおける内容進展の必要性から説明することができる。すなわち、(a)や(b)の第1文の主語「Xが」は、(7)の意味での“焦点”でもある。談話を発展させるためには第2文でこの新情報についての記述を進めることが効率的であり、そのために「(a)主語照応」を頻繁に使用すると考えられる。一方、(c)や(d)の「Xは……Y……」を第1文にする場合は、前節で述べたよう

に、第1文なので旧情報を含むことはあり得ないが、「は」を使用することでXをYより相対的に古い情報として位置付け、次の文からはYについての記述を行って談話が進展させることが多い。そのため、「(d)異主題」が多く観察されると考えられる。

さて、以上の4種の談話の構造と視点の関係を考えるときに参考になるのが次の久野⁷⁾の条件である。

(16) **新主題省略**

(b)「Xが……Y……. Yは……」の「Yは」は、第1文、第2文、ともYの目から見た記述か、XよりもYよりの視点から見た記述である場合に限り、省略できる。

(17) **異主題省略**

(d)「Xは……Y……. Yは……」の「Yは」は、談話の視点が完全にXのそれと合致し、第2文も、Xの目から見たYの記述である場合に限り省略できる。

上の2条件とも、第2文の主題省略のための必要十分条件を述べたものではないが、本論文での焦点の定義を組み合わせると、次のように述べ直すことができる。まず、「Xが」を第1文とする場合については、次の規則を仮定する。各例で下線を付した部分が焦点である。

(18) **「Xが……Y……」型第1文の焦点**

- i) これに続く第2文に対し、Xに対する対立候補Yが存在しないなら、第1文の焦点は「X」である。

(例)夕暮れの動物園、一人の男がゾウののり前を通りかかった。彼は催眠術の分野で非常に優れた才能の持ち主だった。

- ii) この第1文、および第2文ともYに視点がある場合、第1文の焦点は「Y」である。

(例)突然、男がたかしになぐりかかってきた。彼は、必死に防戦しながら、逃げるチャンスがなかった。

- iii) この第1文、および第2文ともXに視点がない場合、第1文の焦点は「Y」である。ただし、ii)より可能性は低い。

iv) そうでなければ、第1文の焦点は「X」である。規則(18)は、既に述べた「(a)主語照応」の方が典型的であるという事実と一致し、特に大きな問題はない。一方、第1文が主題「Xは」を有する場合、久野の条件(17)から(18)に対応する規則を導くと、「(c)反復主題」より「(d)異主題」の方が高頻度であると

* この分類は久野⁷⁾による。ただし、(a)だけは久野の用語と異なる。久野は(a)「主語照応」を「主語を先行詞とする主題省略」と呼んでいる。

いう事実と反する仮説が得られてしまう。そこで、「(c)反復主題」および「(d)異主題」の例文を更に検討してみる。

まず、(c)または(d)に分類されているという点を除いてほとんど構造の等しい談話 (D 1), (D 2) に着目する。それぞれの第2文の解釈は、(D 1-2)では「彼女」=「令子」であり、(D 2-2)では「彼女」=「貴子」である。これらの違いは、次のような第2文の述部にあるだけである。

(D 1-2)' あわてて追いかけた。

(D 2-2)' とてもきれいに着飾っていた。

これらの述部とそれぞれの主題「彼女は」との関係を整理すると、(D 1-2)では「彼女」が動作性の強い述語「追いかける」の文字どおりの動作主 (Agent) になっているのに対し、(D 2-2)では、「彼女」が「着飾る」の Agent となつてはいるものの、文全体としては Agent の動作ではなく状態を述べた文となっている。しかも、(D 2-2)は、その「彼女」の状態を「彼女」自身とは別の視点から記述した文となっている。このことを端的に示しているのが (D 2-2)におけるアスペクト形式素「～ている」の使用である。このことを念頭において(d)に分類される談話を観察する。たとえば、次のような例である。

(D 7) 1. 先日、たかしは令子を病院に見舞った。

2. 思ったより元気だった。

(D 8) 1. 昨日、友達ときれいな花をさがしにいった。

2. でもそれはとうとう見つからなかった。

(D 7-2)では、「元気だった」という形容詞を使って「令子」 (=Experiencer) の状態を「たかし」の視点から記述したものとなっている。また、(D 8-2)は「見つかる」という動詞の Object として第1文の「きれいな花」を記述した文である。このように、(d)に分類される談話の第2文は、いずれも Experiencer や Object の状態記述になっているという共通点がある。

このような状態記述を行う動詞の分類とそれぞれの動詞の格構造との関係を考察したものに文献 12)がある。また井上¹³⁾では、そのような動詞と時の解釈との関係について詳しく考察している。これらの分類で

A類:「ある」、「いる」、形容詞、形容動詞など*

と呼ばれる動詞は、[+状態]という意味素性を有し

* 形容詞の分析はまだ十分でないため、ここではA類としてまとめてある。しかし実際には、形容詞にも主観性の強いものとそうでないものがあり、そのような分類の考慮^{9), 11)}が必要になるとと思われる。

Object の状態を記述する働きがある。つまり、これらの動詞を述部を使用した場合、その Object には視点がないことを示すことになる。この点でA類の動詞と類似の働きをするものとして、通常「～ている」と共に使用する:

B類:「似る」、「そばえる」、「すぐれる」などの動詞がある。また、動詞ではないが、次の (D 9-2) や (D 10-2) などの名詞文も対象の状態を記述するものであり、普通は、その対象に視点があることはない。

(D 9) 1. ある日、ぼくは、小山に囲まれた不思議な場所を見つけた。

2. そこは、小指ほどの人たちが住むコロポックル小国だった。

(D 10) 1. 今朝、たかしはクラスのけんいちから本を返してもらった。

2. 一週間前に貸した鉄道の写真集である。

さらに、[-状態]を素性として Agent をとるA類、B類以外の動詞もアスペクト形式素を伴って、前述の例文

(D 2) 2. 彼女は、とてもきれいに着飾っていた。

のように、[+状態]となつて、その Agent についての状態記述するのに使われる。

これらは、いずれも、状態を記述している対象以外に視点があることを示すものであり、(11)の「正の表現」に対し**負の表現**と呼ぶことができる。負の表現の具体例と視点との関係は、次のようにまとめることができる。

(19) 述部による視点の推定 (負の表現)

- i) 述部の動詞がA類、B類^{13), 12)}のとき、その Object (または Experiencer) は視点ではない。
- ii) 名詞文 (「(Mは) Nだ。」) であるとき、Mに視点は無い。ただし、Mが話し手自身を言及する場合は、Mに視点がある。
- iii) [-状態、+Agent]を素性とする動詞にアスペクト形式素が接続して [+状態] となった場合、その Agent が話し手でない限り、Agent は視点ではない。

視点についての負の表現は、どこに視点があるかを明示するものではないが、“それ以外に視点がある”という情報を与えるもので重要である。同様の機能は、名詞句が担う場合もあるが、ここでは、次の対称詞¹⁰⁾の例をあげるにとどめておく。

(20) 名詞句による視点の推定 (負の表現)

「AのB (対称詞)」という談話登場人物の「B」には視点はない。

(例)次郎の妹は、今年大学を受験した。

以上により、本論文では、視点推定に対する正と負の規則を設定したことになる。これは、視点という概念が相対量であるということの反映でもある。たとえば、規則 (11-i) の授受表現のように、それを使用することで視点がどこにあるかを強く主張する表現もあれば、視点をそれほど明確には主張しない表現も存在する。2章のはじめに述べたように、共感度を自己同一視化の度合として定義するのも本質的に同じことである。また、たとえば、上の規則 (19) に至る議論は、Experiencer より Agent の方が視点に対し強く働くことを暗に含んでいる。実際、Agent を正の視点、Experiencer を負の視点と近似できなくもない。しかし、視点を決定的には左右しないこのような要因を考慮するには、他の要因との相互関係にも注意を払う必要がある。他の要因の例として、ここでは、Xの対立候補Yの有無に着目する。対立候補Yとは、第1文の主語または主題Xに対し、Xと何らかの意味での素性を共有しているために、XとYとのいずれが第2文の主題となっているかが曖昧となる要素である。通常、対立候補Yは、第2文の述部が要求する意味的制約を満たすものであるためにそのような曖昧性を生むことになる。次の例は、第1文の主題「広瀬さんたち」の対立候補が存在しない例である。

(D 11) 1. 広瀬さんたちは、奈良へ遊びに行きました。

2. 奈良の人々の心づかひや新緑の美しい自然に感激しました。

この例では、第1文に第2文の述部に対するXの対立候補Yが存在しないため、焦点はXの「広瀬さん」となる。「広瀬さん」以外が焦点となるためには、そのような要素が現に第1文に存在していなくてはならない。(D 11) の第2文の「感激しました」は Experiencer をとる状態の記述である。しかし、一般に、視点がXにあり、それに続く状態記述の第2文がXの目から見たYの記述となるためには、第1文に対立候補Yが存在する必要がある。以上の点をまとめ、(18) に対応する規則として、次を設定する。

(21) 「(Xは) …… Y ……」型第1文の焦点*

i) これに続く第2文の述語に対し、Xに対する対

立候補Yが第1文に存在しないなら、第1文の焦点は「X」である。

(例) (D 11) 広瀬さんたちは、奈良へ遊びに行きました。奈良の人々の心づかひや新緑の美しい自然に感激しました。

ii) 第1文の視点がXにあり、かつ、第2文が視点Xの目から見た状態の記述である場合 (そのような場合としては、第2文の述部に視点に対する負の表現を使った場合がある)、第1文の焦点は「Y」である。

(例) (D 2) 昨日、令子は街で貴子を見かけた。彼女は、とてもきれいに着飾っていた。

iii) それ以外 (多くの場合、第2文の主題は Agent である) なら、第1文の焦点は「X」である。

(例) (D 1) 昨日、令子は街で貴子を見かけた。彼女は、あわてて追いかけた。

さて、これまでの議論では、第1文の主語または主題「X」に対し、焦点に対する対立候補「Y」が一意に定まっていることを暗黙の前提としてきた。ここで「Y」の選択について簡単に言及する。「X」が動詞「は」もしくは「が」を伴ったものとすれば、「Y」は「X」以外で第1文に出現したすべての名詞句と第1文で記述した事実などがその資格を有しうる。ただし、それらには優先順位があると考えられ、また第2文の述部からも意味的な制約を受ける。優先順位は動詞にも依存し、詳細は今後の課題であるが、記述を完結させるために、仮の規則として次を仮定する。

(22) 「X {は|が} …… Y ……」型第1文でのYの選択

第1文の要素で、第2文の述部の意味的制約を満たす要素がX以外に存在する場合、その要素を対立候補Yとして選択する。Yの選択候補が、複数存在している場合は、次の順に従うものとする。

Agent > Experiencer > Object > Beneficiary > Goal > Source > その他の格要素 > 出来事 > 述部動詞

4. 実験と議論

これまで述べてきた視点および焦点の推定規則に基づいて日本語の照応解決を行うシステムを実現した。システムの記述言語は、KCL (Kyoto Common Lisp) である。解析速度は一談話 (2文) につき3~5秒程度 (SPARCstation 2) であり、現在の辞書登録語数は850語程度である。本システムでは、ローマ字分か

* 「(Xは)」は省略される場合があることを示す。

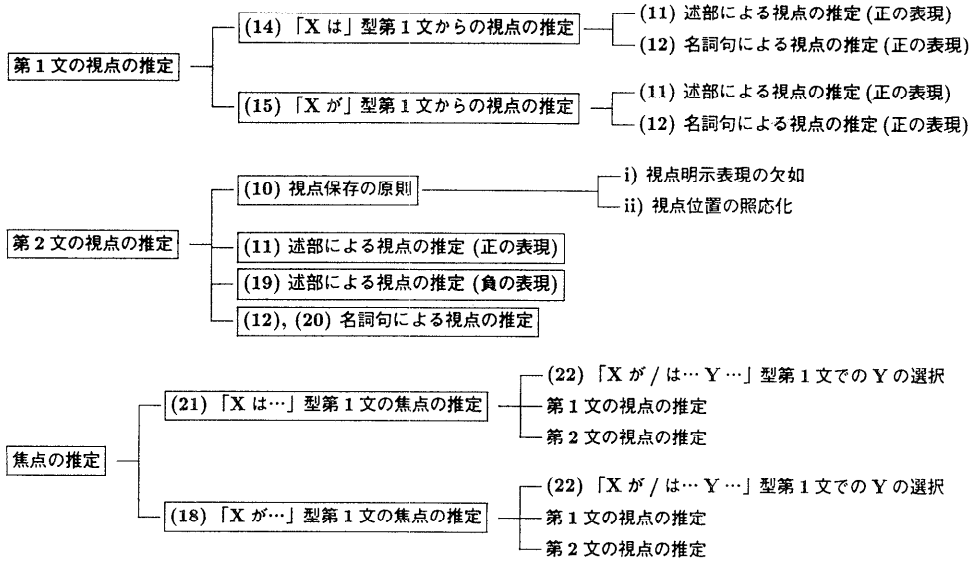


図 1 視点および焦点の推定規則の適用関係
 Fig. 1 Relations of the proposed rules for the prediction of empathy and focus.

```

    CMD >dis
    1ST: 文入力 >haiributoiu ryousiga sirohebiwo
           tasuketeyarimasita /
    2ND: 文入力 >oreini takarano tamawo moraimasita /

    第1文解析結果 (1):(TASUKERU_44)
    第2文解析結果 (1):(MORAU_49)
    -----
    TASUKERU_44 助ける
    (AGENT = (RYOUSHI_40))
    (OBJECT = (SIROHEBI_0))
    RYOUSHI_40 猟師
    (NAMED = (HAIRIBU_0))
    HAIRIBU_0 ハイリブ
    SIROHEBI_0 白蛇
    -----
    MORAU_49 もらう
    (OBJECT = (TAMA_45))
    (CAUSATIVE = (OREI_0))
    TAMA_45 玉
    (OF = (TAKARA_0))
    TAKARA_0 宝
    OREI_0 お礼
    -----
    ガ型第1文 X: RYOUSHI_40
    第1文の視点: RYOUSHI_40
    第2文の視点: RYOUSHI_40
    対立候補 Y: SIROHEBI_0
    第1文の焦点: RYOUSHI_40

    MORAU_49 もらう
    (OBJECT = (TAMA_45))
    (CAUSATIVE = (OREI_0))
    (BENEFICIARY = (RYOUSHI_40))

    CMD >
    
```

図 2 (D4) に対する実行結果. 英小文字はシステムへの入力である.
 Fig. 2 System output for the discourse (D4).

ち書きとローマ字ベタ書きの2通りの入力が可能である。解析は、入力文に対して形態素解析と係り受け解析を行った後、抽出した意味構造を利用して視点、焦点を推定し、照応解決を行う。係り受け解析には文献 14)の主題化の原則に基づくシステムを利用し、可能な係り受け解釈をすべて保持した解析を行う。係り受け解析およびその後の照応解析中に矛盾が生じた結果はその時点で棄却し、解析終了後に残っている結果だけを出力する。図 1 に、このシステムで利用した各推定規則の相互関係を示した。

本システムを用いて実際の談話を入力し照応解決を行った際の実行結果を示し、簡単に説明する (図 2, 図 3 参照)。例として、まず、

- (D 4) 1. ハイリブという猟師が白蛇を助けてやりました。
 2. そして、宝の玉をお礼にもらいました。

を考える。この例の第 1 文は「Xが」型であり、次のような順序で照応解析が進行する。

- 最初に、第 1 文の視点を推定するために (15) を適用する。述部に「てやる」が存在するため、(11-i-2) により、「ハイリブという猟師」に視点があることになる。
- 次に、「視点保存の原則」の (10-ii) により第 2 文の視点も同じ「ハイリブとい


```

CMD >dis
1ST: 文入力 >senzitu takasiha reikowo byouinni mimatta /
2ND: 文入力 >omottayori genkidatta /

第1文解析結果(1):(MIMAU_24)
第2文解析結果(1):(GENKI_25)
-----
MIMAU_24 見舞う
  (AGENT = (TAKASI_0))
  (OBJECT = (REIKO_0))
  (LOCATIVE = (BYOUIN_0))
  (TIME = (SENZITU_0))
TAKASI_0 たかし
REIKO_0 令子
BYOUIN_0 病院
SENZITU_0 先日
-----
GENKI_25 元気
  (CF = (OMOU_10))
OMOU_10 思う
-----
ハ型第1文 X: TAKASI_0
第1文の視点: TAKASI_0
第2文の視点: TAKASI_0
対立候補 Y: REIKO_0
第1文の焦点: REIKO_0

GENKI_25 元気
  (CF = (OMOU_10))
  (EXPERIENCER = (REIKO_0))
CMD >

```

図3 (D7) に対する実行結果

Fig. 3 System output for the discourse (D7).

う猟師」となる。

- (22)により、第2文の述部に対するXの対立候補「Y」として「白蛇」を選択する。
- 第1文の焦点を(18)によって推定すると、(18-iv)により、焦点は「ハイリブという猟師」となる。

すなわち、第2文で照応があるとすると、まず、「ハイリブという猟師」の可能性が最も高いということになる。実際、第2文には主題の省略があり、その先行詞は「ハイリブという猟師」である。つまり、「ハイリブという猟師が宝の玉をお礼にもらった。」という解釈が得られる。「誰からもらったのか」という点については本論文の議論だけでは定めることができないが、残っている可能性は「白蛇」しかないため、その決定は容易であろう。

次に別の例として、

- (D 7) 1. 先日、たかしは令子を病院に見舞った。
2. 思ったより元気だった。

の解析の概略を示す。

- (14-i), (12-ii)により、視点は「たかし」になる。
- (22)により、第2文の述部に対するXの対立候

補「Y」に「令子」が選択される。

- (19-i)にいう負の表現「元気だった」が存在するので、(21-ii)から、第1文の焦点はYの「令子」となる。

よって、第2文で省略されている主題は「令子」であり、「令子は思ったより元気だった」という解釈が得られる。このほか、童話や小学校の国語の教科書から採用した談話例で本システムをテストし、多くの例で妥当な解釈が得られることを確認した。

最後に、本論文のモデル全体についての考察を前提(1)~(3)との関係で述べる。まず、照応解決の代表的モデルである Centering と本論文のモデルの結論とも言うべき規則(18)および(21)とを比較してみる。最初に、「Xが……Y……」型第1文の焦点の推定規則である(18)を考える。ここでは、大まかに言って、ii), iii)で視点が調べられ、二つの文で視点が共有されている場合にはその視点自身が焦点と認められる。そうでなければ、主題のXが視点となる。これは、規則(18)が想定している

談話を考える限り、Kameyama⁵⁾の制約の適用と本質的に全く等しい。つまり二つのモデルに実質的な差はなく、この点で、アプローチの異なるモデルが同じ結論を得ていることは興味深い。同時に、本論文ほど限定の強くない Centering の方が一般的であると言える。しかし、さらに規則(21)を比較すると事情は少し変化する。(21-ii)は、Kameyama⁵⁾の用語によれば、視点がそこにあることを示す“nonIDENT”という属性(property)の共有に基づく照応である。

Kameyama⁵⁾はその場合を優先順位の低いものとしているのに対し、本論文の(21)ではこれを頻度の高いものとしている。しかも、この点は第1文に依存し、規則(18)と(21)との場合分けは、第1文が主題を持つかどうかにかかっている。Centeringの枠組で主題を明確に考慮しているのは文献10)であるが、単に属性間の線形な優先順序に主題を追加しただけであり、やはり、本論文の(18)と(21)との場合分けのような現象に適切に対処することはできない。

本論文の規則(18)と(21)との場合分けは、多くの談話例を現実を観察した結果に基づいているが、より端的には、久野の条件(16)と(17)とを肯定的に利用した結果でもある。久野の条件を照応解決で利用する際に考慮すべき項目を挙げると、まず、照応化の対象とし

て第2文の主題しか問題にできないことが指摘できる。これは、本論文の立場では原則(9)によって規定した点に対応する。第2文の主題以外の照応に対する拡張可能性については、このような規定とは別に議論する必要がある。次に、談話の視点が第1文と第2文とで明らかになっていなければならないという点を指摘することができる。既に2章で述べたように、視点とは談話の話し手が自分と同一視しようとしている対象であり、その暗黙値は言うまでもなく話し手自身である。“話し手”が非常に緩い制約で談話の開始の文ですらゼロ化されるのもこのためである。しかし、日本語のこのような特性をより明確にしようとすると、第1文における「は」、「が」の使用の別や「は」の用法の第1文、第2文での違いにまで踏み込む必要が出てくる。従来の「は」の用法の説明の多くは、このような依存性を暗に前提としたものであっても、それを明確に分離せずに用法をすべて羅列する傾向が強かった。照応現象そのものを深く議論しようとする場合には、むしろ、対象となる談話の性格を限定した方が議論の焦点を集中しやすい⁹⁾。日本語談話のこのような特徴を最も端的に表すものが、談話の開始の2文であり、これが前提の(2)である。さらに、第1文における「は」の用法を制限するための制約が前提の(1)である。結局、前提(1)にいう談話の開始の2文のみを対象としたときさえ、そこに含まれる照応現象は Centering が想定しているほど単純ではない、というのが本論文の結論である。むしろ、対象を(1)~(3)のように限定することで、「は」、「が」の使用による談話の4分類が可能となり、単純とは言えない現象にある程度見通しのよい議論を提供することができると考えられる。

5. おわりに

本論文では、助詞「は」、「が」の使用により談話を4種類に分類し、視点抽出の具体的方法とそれを利用した照応解決の手法を提案した。さらに、それを計算機上のシステムとして実現し、その有効性を確認した。

本論文の手法は、1章の(5)に述べたように、照応の対象となりやすい要素を実際の照応の有無とはある程度独立に明らかにするという方針に基づいている。特に、照応の可能性の最も高い要素を焦点として抽出するという実現法を採っているため、複数の照応がある場合には直には対処できない。さらに、処理の対象を談話の開始の2文に限定しているため、談話中の一般の隣接文間の照応を扱う場合にもモデルの拡張が

必要となる。これらに対処するには、Centering の枠組にある、属性あるいは素性の共有という考えを積極的に採用することが必要となろう。そのような属性としては、本論文のモデルで利用している**視点**、**主題**および**意味的な格**に加えて、主語、目的語などの**表層格**の考慮も必要になるはずである。しかし、その場合であっても、これらの概念は全く次元の異なるものであることに注意が必要である。本論文のモデルは、Centering の制約一全く次元の異なるこれらの概念に単純な順序関係を想定することの妥当性を検証する上でも役に立つものである。

謝辞 文献等を指摘下さった横浜国立大学電子情報工学科中川裕志助教授、武藤伸明氏ならびに査読者に深謝する。

参考文献

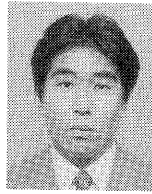
- 1) 清水一澄, 横尾英俊: 視点と焦点の関係に基づく日本語談話の照応解決, 情報処理学会研究報告, Vol. 93, No. 79, pp. 13-20 (1993).
- 2) 山梨正明: 推論と照応, くろしお出版, 東京 (1992).
- 3) 中川裕志, 西澤信一郎: 日本語の理由-行為の順接複文におけるゼロ代名詞照応, 情報処理学会論文誌, Vol. 35, No. 10, pp. 2038-2045 (1994).
- 4) Sidner, C.L.: Focusing in the Comprehension of Definite Anaphora, Brady, M. and Berwick, R.C. (eds.), *Computational Models of Discourse*, pp. 267-330, MIT Press, Cambridge, MA (1983).
- 5) Kameyama, M.: Japanese Zero Pronominal Binding: Where Syntax and Discourse Meet, Poster, W. (ed.), *Japanese Syntax*, pp. 47-73, CSLI, Stanford, CA (1988).
- 6) 久野 暉: 日本文法研究, 大修館書店, 東京 (1973).
- 7) 久野 暉: 談話の文法, 大修館書店, 東京 (1978).
- 8) Kuno, S.: *Functional Syntax*, University of Chicago Press, Chicago, IL (1987).
- 9) 山村 毅, 大西 昇, 杉江 昇: 日本語の照応問題解決システム, 電子情報通信学会論文誌 D-II, Vol. J73-D-II, No. 6, pp. 887-896 (1990).
- 10) Walker, M., Iida, M. and Cote, S.: Centering in Japanese Discourse, *COLING-90*, Suppl., pp. 1-6 (1990).
- 11) 武藤伸明, 中川裕志: 感覚述語におけるゼロ目的語の推定, 情報処理学会論文誌, Vol. 34, No. 3, pp. 409-417 (1993).
- 12) 井上和子(編): 日本文法小事典, 大修館書店, 東京 (1989).
- 13) 井上和子: 変形文法と日本語(下), 大修館書店,

東京 (1976).

- 14) 横尾英俊, 田島恵美子: 日本語係り受け解析のための主題化にかかわる基礎原則, 電子情報通信学会論文誌 D-II, Vol. J75-D-II, No. 10, pp. 1755-1759 (1992).

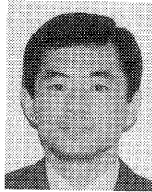
(平成6年3月31日受付)

(平成6年12月5日採録)



清水 一澄 (正会員)

1992年群馬大学工学部情報工学科卒業. 1994年同大学院博士前期課程修了. 同年シャープ株式会社入社. 現在, 同社情報システム事業本部所属.



横尾 英俊 (正会員)

1978年東京大学工学部計数工学科卒業. 1980年同大学院情報工学専門課程修士課程修了. 同年山形大学工学部電気工学科助手. 1989年群馬大学工学部情報工学科勤務. 現在, 同学科助教授. 工学博士. データ圧縮, 自然言語理解などの研究に従事. 電子情報通信学会会員.