

語用論的・意味論的制約を用いた日本語ゼロ代名詞の文内照応解析

6R-3

中岩浩巳 池原悟

NTT コミュニケーション科学研究所

1. はじめに

自然言語では通常、相手に容易に判断できる要素は、文章上表現しない場合が多い。よって、例えば機械翻訳システムにおいては、原言語での省略要素が、目的言語で必須要素になる場合、省略要素の同定が必要となる。特に日英機械翻訳システムにおいては、日本語の格要素が省略される傾向が強いのに対し、英語では訳出上必須要素となるため、この省略された格要素（ゼロ代名詞と呼ばれる）の照応解析技術は重要となる。

従来からこのゼロ代名詞の照応解析に関して、様々な手法が提案されている[1][2][3][4]。これらの手法を、翻訳対象分野を限定しない実用的機械翻訳システムに応用することを考えると、解析精度や必要となる知識量が膨大となる面で、そのままでは実現不可能である。我々は、知識量の爆発を避けるための手段として、用言のもつ意味を分類し、その語のもつ代表的属性値によって語と語や文と文の意味的関係を決定して、他の文に現われる照応要素を決定する手法を提案した[5]。また、照応要素が同一文中や文章の他の文中には現われないゼロ代名詞に対して語用論的意味論的制約を用いた照応解析を行なう手法を提案した[6][7]。

本稿では、照応要素が同一文中に現われるゼロ代名詞に対して、接続語のタイプや用言意味属性等の語用論的・意味論的制約を用いた汎用的な照応解析を行なう手法を提案する。

2. MT評価用例文でのゼロ代名詞の出現傾向

照応要素が同一文章中あらわれるゼロ代名詞の傾向を掴むために、本章では独立した文（文間文脈情報が得られない文）におけるゼロ代名詞を調査した。調査対象は、日英機械翻訳システム評価用例文3718文[8]である。個々の文には模範となる英訳が付与されており、そのほとんどの文は文脈の情報無しに翻訳が可能である（3718文中3704文）ため、個々の文を日英機械翻訳システムで翻訳し、予め用意された英訳と比較することでシステムの翻訳機能の評価が行なえる。また、個々の例文は自然な日本語文であり広範囲な表現が含まれているため、これらの例文におけるゼロ代名詞とその照応要素の出現傾向を調査することによって、同一文内にも照応要素がある場合や、文中に現われない照応要素の傾向を把握することが可能となる。

この試験文に対して照応解析が必要となるゼロ代名詞とその照応要素の出現傾向を調査した結果を表1に示す。照応要素の出現場所は、大きく、同一文内に存在する場合と、同一文内に存在しない場合に分かれる。調査結果によれば、全ゼロ代名詞512件に対して照応要素が同一文中に現われるゼロ代名詞が139件（27%）であった。また、照応要素が文中に存在しない場合が373件（73%）存在する。この373件の詳細については既に報告しているので[6][7]、ここでは照応要素が同一文中に現われるゼロ代名詞の詳細について述べる。

照応要素が同一文中に現われるゼロ代名詞のうちでは、

ガ格がゼロ代名詞化され照応要素がハ格の要素となる場合が102件と最も多い。このなかで、ゼロ代名詞が照応要素より文章中の前の部分に存在する後方照応表現が8件存在した。この現象は、助詞の種類に基づく前方照応解析手法では解析することが出来ず、接続語のタイプ等に基づく照応解析が必要となる。

ゼロ代名詞化されたのと同じ格の要素が照応要素となる（例えば、ガ格がゼロ代名詞化され同一文内のガ格が照応要素となる）場合が、10件存在する。これは、接続語のタイプにより、同一文内の格要素が共有できるかが決まるという特質を用いることで解析可能となることが予想される。

また、照応要素が同一文中に存在するゼロ代名詞の中で、埋め込み文又は引用文中の格要素がゼロ代名詞化されている場合が9件、照応要素が埋め込み文又は引用文中に存在する場合が4件あった。これらのゼロ代名詞を正しく解析するためには、埋め込み文や引用文と同一文内のそれ以外の表現との意味的関係を用言意味属性や様相表現、埋め込み文が修飾する名詞のタイプ等の情報を用いて決定することが必要となる。

表1. ゼロ代名詞とその照応要素の出現頻度

(調査対象文：日英機械翻訳システム評価用例文3718文、照応解析を要するゼロ代名詞が現われる文は463文、512件)

照応要素の出現場所	同一文内										文章中になし				小計(件)
	は	が	を	に	他	受身	I	you	人	it	他	文章外			
												引用文内	埋込文内		
は	1	0	0	0	0	0	5	0	0	0	2	0	0	8	
が	102	2	6	2	0	1	7	51	9	28	23	50	3	444	
埋込文内	3	0	1	0	0	0	0	15	0	0	2	0	0	21	
引用文内	4	0	0	0	0	0	0	0	0	0	0	0	0	4	
を	3	0	0	0	3	1	0	0	0	0	11	0	0	18	
埋込文内	1	0	0	0	0	0	0	0	0	0	0	0	0	1	
に	1	0	0	0	0	0	0	2	2	5	0	0	2	12	
他	0	0	0	0	1	0	0	0	1	1	0	1	0	4	
小計(件)	139										373				512

3. ゼロ代名詞の同一文内照応解析

2章で得られた結果を元に、照応要素が同一文中に現われるゼロ代名詞の解析手法について提案する。

3.1 助詞の種類を用いた文内照応解析

日本語ゼロ代名詞の文章中に現われる照応要素を決定する解析手法としては、助詞のタイプや共感動詞の有無に基づき格要素のCenterをランク付けし、文間の話題の継承性を認定することによってゼロ代名詞の照応要素を決定するcenteringアルゴリズムが知られている[1][2]。この手法は、文中に含まれる個々の複文を単文に分割して適用することによって、文内照応にも適用可能である。例えば、

(1)彼は方程式を解いて答えを出した。
 という文では「文を「彼は方程式を解く」と「(φ)が答えを出す」に分割し、動詞「出す」のガ格のゼロ代名詞の照応要素として助詞「で」が主語化された「彼」が認定される。本手法は、アルゴリズムが極めて簡単であるため、実現が容易であるが、解析精度の面では不十分である。例えば、本手法は前方照応指示を解析対象としているため、次に様な後方照応指示表現は解析不可能である。

(2) (φ)が 繩を枝から枝にかけて、子供達は遊んだ。

この例ではゼロ代名詞の照応要素は後半の用言のハ格である「子供達」となるが、ゼロ代名詞が照応要素より前の単文にあるため、このアルゴリズムでは解析不可能となる。

3. 2 論用論的意味論的制約を用いた文内照応解析
本節では、2章で調査した例文を詳細に検討することにより得られた、接続語、用言意味属性、様相表現の3種類による文内照応要素を決定するための語用論的・意味論的制約について述べる。

3. 2. 1 接続語による制約

接続語は、ゼロ代名詞の文内照応要素を決定するうえで最も強力な制約となる。南[9]や田窪[10]らが提案しているとおり、日本語の接続語は、格要素の影響範囲を制限するものがある。例えば、南は、日本語接続詞を格要素の影響範囲に応じてA,B,Cの3種類に分類した。この分類によると、A類の接続語を伴う複文において、この接続語をはさんだ片方の単文のガ格がゼロ代名詞化されもう片方の単文にガ格が存在する場合には、このゼロ代名詞の照応要素はもう片方の単文のガ格となる。このような接続詞の種類による格要素の特徴を照応要素の決定に活用することができる。

3. 2. 2 用言意味属性[11]による制約

用言意味属性による制約は大きく以下の2種類に分かれる。

(a) 用言意味属性による制約

ゼロ代名詞又はその照応要素が埋め込み文又は引用文中にある場合の文内照応解析は、埋め込み文又は引用文の表現と文内のその他の表現との意味的關係を認定することが必要となる。この意味解析に、文中の用言の用言意味属性が有効となる。接続語のタイプで照応要素が決定できない場合でも、用言の意味属性を利用することで用言間の意味的關係が認定でき、文内照応解析が可能となる。

(b) 用言意味属性と接続語による制約

3. 2. 1で示した接続語のタイプによる格の共有による制約は、文内照応解析に極めて有効である。しかし、接続語の種類によっては1種類の接続語が複数の接続語タイプの曖昧性を持つ場合がある。例えば、接続語「て」は、接続語の前後の文の關係に関して、動作の様態を示すAタイプの意味、時間關係や原因を示すBタイプの意味、並列表現を示すCタイプの意味がある。このような接続語に対しては、接続語の前後の用言意味属性、様相表現の種類の共起によって接続語の意味を特定し接続語のタイプを決定することが出来る。

3. 2. 3 様相表現による制約

様相表現は、ゼロ代名詞の文章外照応要素を決定するうえで強い制約となる[6][7]。例えば、ガ格がゼロ代名詞化している場合には、様相表現「～したい(φ want to～)」(希望)を伴うと、照応要素は「I」になり、様相表現「～してはいけない(φ must not～)」(禁止)を伴うと、照応要素は「you」になると言える。このような特長は、文内に「you」や「I」が存在する場合の文内照応解析にも有効と考えられる。

3. 3 アルゴリズム

3. 1や3. 2で示した条件を元に、照応要素が同一文内に現われるゼロ代名詞の解析アルゴリズムについて提案する。前述の条件をアルゴリズム化するには、照応要素が文章中に現われない場合や、他の文に現われる場合の解析精度も考慮に入れて、全体的にゼロ代名詞の解析精度が良くなるように実現する必要がある。処理アルゴリズムは、以下のとおりである。なお、各ステップで照応要素を決定する際には、アルゴリズム中に記した条件だけではなく、用言がゼロ代名詞に課す意味的制約を満たすかも検証する。

<Step-1> ゼロ代名詞を検出する。

もし検出されれば、現在解析中の文のタイプによって処理を分類する。複文・重文の場合はStep-2へ、単文の場合はStep-3へ。

<Step-2> 照応解析を下記の順で行う。

1)用言意味属性、様相表現および接続語の種類による文内照応解析(条件:3.2.2(b), 3.2.3)

2)接続語のタイプによる文内照応解析(条件:3.2.1)

照応要素が決定すれば解析終了,決定しなければStep-3へ。

<Step-3> 現在解析中の文に埋込文や引用文が含まれている場合、用言意味属性の制約を用いた文内照応解析。

(条件:3.2.2(a))

照応要素が決定すれば解析終了,決定しなければStep-4へ。

<Step-4>他の文に照応格要素が存在するか調査する。

(例えば、[5]の手法で調査する)

照応要素が決定すれば解析終了,決定しなければStep-5へ。

<Step-5> ゼロ代名詞が支配する用言の用言意味属性、様相表現および接続語の種類による文章外照応解析[6][7]。

照応要素が決定すれば解析終了。

<Step-6> 照応要素が決定できない場合、用言がゼロ代名詞に課す意味的制約により照応要素を推測。また、受け身変形可能な場合は受け身化し解析終了。

4. 評価

表1の文のうち、文内照応解析が必要となるゼロ代名詞139件を詳細に調査して作成したルールを用いて、文章外照応解析が可能となるゼロ代名詞(ガ格-I or we, ガ格-you, ガ格-人, ガ格-it, ニ格-you)の5種類,175件)および、受け身変形することにより照応解析が不要となるゼロ代名詞173件を含めた計487件を対象に、照応解析を行なったところ、再現率98%,適合率100%の精度で正しく文章外の照応要素が推測された。Centeringアルゴリズム[1][2]による結果は再現率74%,適合率89%であることから、本手法の有効性が確認できた。

5. おわりに

本稿では、用言意味属性や表層表現を用いた日本語ゼロ代名詞の文内照応解析手法について提案した。今後は、ルールの詳細化を行ない汎用性の検証を行なっていく。

<参考文献>

- [1]Kameyama,M: A Property-sharing Constraint in Centering, Proc of ACL, (1986).
- [2] Walker, M. et al: Centering in Japanese Discourse, Proc of COLING'90 (1990).
- [3] Yoshimoto, K.: Identifying Zero Pronouns in Japanese Dialogue, Proc of COLING'88 (1988).
- [4] 堂坂: 語用論的條件の解釈に基づく日本語ゼロ代名詞の指示対象同定, 情処論, Vol.35 No.5(1994).
- [5] 中岩,池原: 日英翻訳システムにおける用言意味属性を用いたゼロ代名詞照応解析, 情処論, Vol.34 No.8 (1993).
- [6] 中岩,白井,池原: 用言意味属性を用いた日本語ゼロ代名詞の文章外照応解析, 情処49回全大,(1994).
- [7] Nakaiwa, H. et al.: Extrasyntactical Resolution of Japanese Zero Pronouns using Semantic and Pragmatic Constraints, Proc. of AAAI 1995 Spring Symposium Series, Empirical Methods in Discourse Interpretation and Generation (1995).
- [8] 池原,白井: 日英機械翻訳機能試験項目の体系化, 電子情報通信学会研究会, Vol.NLC-90 No.43, pp.17-24 (1990).
- [9] 南: 現代日本語の構造, 大修館書店 (1974).
- [10] 田窪: 統語構造と文脈情報, 日本語学, Vol.6No.5 (1987).
- [11] Nakaiwa, H. et al.: A System of Verbal Semantic Attributes Focused on the Syntactic Correspondence between Japanese and English, Proc.of COLING'94 (1994).