

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

4G-10

中岩浩巳 白井諭 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（読み手もしくは聞き手）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。この現象は、機械翻訳システムや対話処理システム等の自然言語処理システムにおいて、大きな問題となる。例えば、機械翻訳システムにおいては、原言語では陽に示されていない要素が、目的言語で必須要素になる場合、陽に示されていない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。KameyamaやWalkerらは、中心化アルゴリズムに基づき助詞の種類や共感動詞の有無により文章中に現われる照応要素を決定する手法を提案した[1], [2]。また、Yoshimotoは対話文に対して文章中にあらわれる照応要素については主題をベースとして照応要素を同定し、文章中に現われないゼロ代名詞については敬語表現やspeech actに基づき照応要素を同定する手法を提案した[3]。堂坂は、日本語対話における対話登場人物間の待遇関係、話者の視点、情報のなわばりに関わる言語外情報の発話環境を用いて、ゼロ代名詞が照応する対話登場人物を同定するモデルを提案した[4]。林らは、マニュアル文章中の日本語受動文が推敲の観点から能動態化が可能であるかを、能動態化した際の主語となるべき要素を堂坂の手法をベースに同定することによって判断する手法を提案した[5]。

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應用することを考えると、解析精度の面や対象とする言語現象が限られる点、また、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て、その語のもつ代表的属性値によって、語と語や文との関係を決定して、同一文や文章中の他の文内に現われる照応要素を決定する手法を提案した[6]。

本稿では、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、用言意味属性を用いた照応解析を行なう手法を提案する。用言意味属性を用いることによって、対象文を限定することなしに文章外の照応要素が決定できる。

2. 日英機械翻訳システム評価用例文でのゼロ代名詞の出現傾向

照応要素が文章中にない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[7]である。この評価用例文は、日本語の性質と表現の種類、及び日本語と英語との相違に基づき体系化された約500種類の試験項目を評価するために実用文中の表現を抽出して作成された日本語例文

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。同一文内に存在しない場合におけるゼロ代名詞は、その訳出方法別に、以下のように分けた。

- ・受け身変形化しゼロ代名詞となる要素を訳出させない場合
- ・Iやweのように筆者や話者本人（I）もしくは本人を含む集団（we）が照応要素となる場合
- ・youのように読者や聞き手が照応要素となる場合
- ・誰かは特定できないが人間が照応要素となる場合
- ・itを生成すべき場合
- ・その他の特定の要素が照応要素となる場合。

調査結果によれば、全ゼロ代名詞536件に対して照応要素が文中に現われないゼロ代名詞が396件（74%）と割合が高い。そのうちで、受け身変形することによってゼロ代名詞の照応要素を認定せずに生成することが可能となる場合は174件（32%）にとどまる。よって残りの222件（41%）のゼロ代名詞はなんらかの手段で文章中には現われない照応要素を認定する必要がある。このなかで、ゼロ代名詞が主語でI又はweが照応要素となる場合78件（15%）を詳細に見てみると、ゼロ代名詞を支配するのは、「～したい」のような希望の様相を伴う表現や、「～しよう」のような勧誘の様相を伴う表現、「思う」のような思考動作を示す用言を伴う表現等となっている。また、ゼロ代名詞が主語でyouが照応要素となる場合27件（5%）を詳細に見てみると、ゼロ代名詞を支配するのは、「～すべきだ」のような義務の様相を伴う表現や、「～してはならない」のような禁止の様相を伴う表現等となっている。このように、照応要素が文章中に現われない

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、調査対象文中でゼロ代名詞を含む文は2718文中463文）

照応要素の出現場所	同一文内					文章になし						小計 [件]	
	は	が	を	に	他	受身	I we	you	人間	it	他		
が	109	7	1	0	0	7	151	78	27	21	65	3	469
埋込文内	2	2	0	0	1	0	15	0	0	2	0	0	22
を	3	0	0	3	1	0	0	0	0	0	12	0	19
に	2	0	0	0	0	0	2	2	5	0	0	2	13
他	0	1	0	1	0	0	6	1	1	0	3	0	13
小計 [件]	140					396						536	

Extrasentential Anaphora Resolution of Japanese Zero Pronouns using Verbal Semantic Attributes.
Hiromi NAKAIWA, Satoshi SHIRAI and Satoru IKEHARA
NTT Communication Science Laboratories

ゼロ代名詞を解析するためには、様相や用言意味属性を用いることが有効となる。

3. 用言意味属性

本手法で用いる用言意味属性の体系を図1に示す[8]。この用言意味属性は、我々の日英翻訳システムALT-J/Eの日英構造変換辞書の各日英パターン対(約15,000対)に付与している。構文的特徴の大きく異なる日本語と英語の対訳関係にある表現にたいして用言意味属性を付与することにより、日本語における用言の多義が絞り込める。

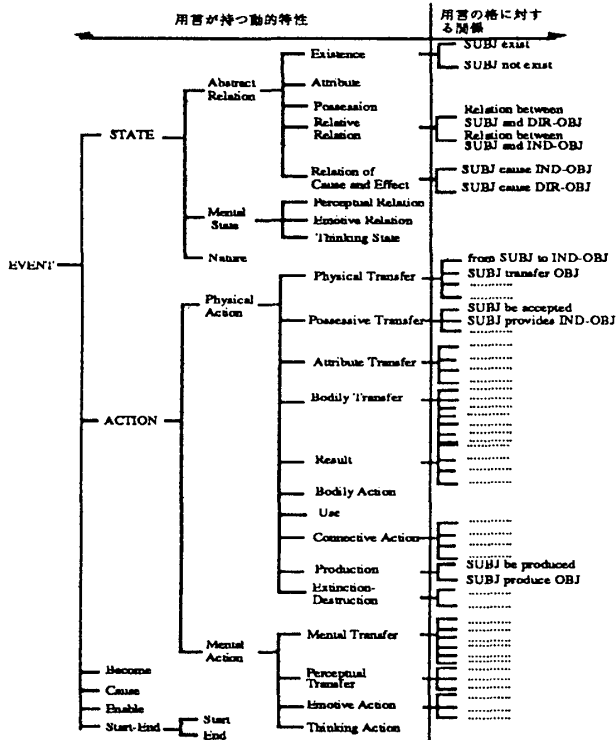


図1 用言意味属性体系

4. ゼロ代名詞の文章外照応解析

2章で得られた結果を元に、照応要素が文章外に現われるゼロ代名詞の解析手法について提案する。ここでは、日英機械翻訳システムで実現した処理例について説明する。よって、ここで照応解析を行なうゼロ代名詞は、英訳するうえで必須要素となるゼロ代名詞のみとなる。処理アルゴリズムは、以下のとおりである。

(ステップ1)

ゼロ代名詞を検出する。

もし検出されれば、同一文内および他の文に照応格要素が存在するか調査する。(例えば、[6]の手法で調査する)存在し、照応解析が終了すれば解析終了。

(ステップ2)

同一文内および他の文に照応格要素が存在しなければ、ゼロ代名詞が支配する用言の用言意味属性および様相表現の種類[9]によって、文章外照応解析を行なう。照応解析の条件は表2に示す通りである。文章外照応要素を決定する際には、この条件のみではなく、用言がゼロ代名詞に課す意味的制約をも満たすかも検証する。

照応要素が決定すれば解析終了。

(ステップ3)

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。

表2 文章外照応要素の推測条件

ゼロ代名詞出現場所	照応要素	条件	考察
が格	I or we	様相:希望(~したい)	話/著者が希望する
		様相:三人称希望,使役(~してほしい)	話/著者が話/読者に希望する
		様相:勧誘(~しましょう)	勧誘するのは話/著者
you	you	用言:Action配下の属性+様相:丁寧(~します)	発話者の待遇関係による
		様相:禁止(~してはいけない)	話/著者が話/読者の構造を禁止する
人(I, we, you,...)	人(I, we, you,...)	用言:Action配下の属性話+様相:義務(~べき)	話/著者が話/読者の活動を義務化
		用言:Bodily Action Thinking Action Emotive Action Emotive State Bodily Transfer	人しか動作しえない用言、感情をしめす用言が出現すると文脈がないと人の解釈がもっともらしい
it	it	用言:名詞述語文で名詞の意味が抽象	抽象概念の代名詞はitが適切
		用言:Attribute & Perceptual State	寒い,暑いなど気候を示す語
に格	you	様相:三人称希望,使役(~してほしい)	話/著者が話/読者に希望する

5. 評価

表1の文の内、表2の条件で解析可能となるゼロ代名詞(が格-I or we, が格-you, が格-人, が格-it, に格-you)の5種類、196件を対象に4章の手法で照応解析を行なったところ、85.5%の精度で正しく文章外の照応要素が推測された。これにより本手法の有効性が確認できた。

6. 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章外照応解析手法について提案した。今後は、表2のルールの詳細化を行ない汎用性の検証を行なうと共に、文内照応解析に関しても検討していく。

<参考文献>

- [1] Kameyama, M. : A Property-sharing Constraint in Centering, Proc of ACL, pp.200-206 (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 pp.779-784 (1988).
- [4] 堂坂：語用論的条件の解釈に基づく日本語ゼロ代名詞の指示対象同定, 情報処理学会論文誌, Vol.35 No.5, pp.768-778 (1994).
- [5] 林, 千葉：日本語受動化可否判定アルゴリズムの検討, 情報処理学会論文誌, Vol.31 No.10, pp.1438-1443 (1990).
- [6] 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, 情報処理学会論文誌, Vol.34 No.8, pp.1705-1715 (1993).
- [7] 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, 言語理解とコミュニケーション研究会, Vol.NLC-90 No.43, pp.17-24 (1990).
- [8] Nakaiwa, H. et.al : A System of Verbal Semantic Attributes Focused on the Syntactic Correspondence between Japanese and English, Proc.of COLING'94 (1994).
- [9] 河合：日英翻訳システムALT-J/Eにおける様相, 時制の処理, 情報処理学会第34回全国大会, pp.1245 (1987).