

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

中 岩 浩 巳[†] 池 原 悟[†]

日本語テキスト中に現れるゼロ代名詞に対する、用言意味属性を用いた照応解析手法について提案する。本手法は、用言が持つ意味的属性に着目し、ゼロ代名詞が支配される用言の意味属性と、それと照応する格要素候補が支配される用言の意味属性の関係から、文脈を調べ、照応格要素を決定するものである。また本論文では、口英言語間の特性の違いから、日英機械翻訳における照応解析の必要条件を検討し、本手法を適用する際のゼロ代名詞の検出条件や具体的な変換方法を明らかにする。本手法を日英翻訳システム ALT-J/E 上に実現して、新聞記事を対象に、2種類の性能評価を行った。まず、照応解析ルールを整備した標本実験では、機械翻訳の訳出に必要なゼロ代名詞の大半(95%)が正しく照応解析されることがわかった。次に、上記実験で使ったルールをそのまま適用した別の標本の実験では、本手法が適用対象とするゼロ代名詞については83%、本手法の適用対象外のゼロ代名詞も含む場合でも73%の照応要素が正しく認定された。さらに、若干の不足ルールを整備した後は、適用対象内の正解率は95%に向上した。以上の結果、機械翻訳における本手法の効果が実証され、省略された格要素の大半が復元できるという見通しを得た。

Zero Pronoun Resolution in a Japanese to English Machine Translation System using Verbal Semantic Attributes

HIROMI NAKAIWA[†] and SATORU IKEHARA[†]

A method of resolution of zero pronouns in Japanese language texts using the verbal semantic attributes is suggested. This method examines the context from the relationship between the semantic attributes of verbs governing zero pronouns and the semantic attributes of verbs governing their referents. The semantic attributes of verbs are created using two different viewpoints: dynamic characteristics of verbs and the relationship of verbs to cases. It is shown that, using this method to translate newspaper articles, the major portion (95%) of resolution of zero pronouns necessary for machine translation can be achieved using only linguistic knowledge. Factors to be given special attention when incorporating this method into a machine translation system are examined, together with suggested conditions for the detection of zero pronouns and methods for their conversion. This study considers four factors that are important when implementing this method in a Japanese to English machine translation system: the difference in conception and the difference in case frame patterns between Japanese and English expressions, restrictions by voice and restrictions by translation structure. Implementation of the method leads to a viable method for resolution of zero pronouns in a practical machine translation system.

1. はじめに

通常自然言語では、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¹⁾。このような省略現象にはさまざまなものがあるが、翻訳システムで特に問題となるのは、原言語では陽に示されていない要素が、目的言語では必須要素となる場合である。特に日本語の格要素(特に主語と目的語)は、読者が容易に推定できる範囲で、省略される傾向が強いのにに対し、英語では訳出上必須要素となる。したがって、日

英翻訳システムでは日本語で省略されている格要素(ゼロ代名詞と呼ばれる)の照応要素を同定し、英語表現に訳出することが必要であり、ゼロ代名詞型の照応解析技術が特に重要となる。

従来から、このゼロ代名詞の照応問題に対しては、さまざまな方法が提案されている。Groszらは、中心化アルゴリズム(centering algorithm)を用いた定名詞句の照応手法を提案した²⁾。Kameyamaはこの考えを拡張し、属性共有制約(property sharing constraints)を導入して、日本語ゼロ代名詞の照応解析に適用した³⁾。この手法では、助詞の種類や共感動詞(empathy-loaded verbs)の有無により、談話セグメ

[†] NTT 情報通信網研究所
NTT Network Information Systems Laboratories

ントのフォーカスの優先順位を制御し、それをもとにゼロ代名詞の照応解析を行っている。また、Yoshimotoは、対話文に対して主題をベースとしたゼロ代名詞照応解析法を提案した⁴⁾。この方法は、文章の格が助詞の型によって決まるという日本語の特徴に着目したもので、助詞“は”を伴う格要素および、特殊な構文パターンに支配された表現によって主題化された格要素をゼロ代名詞の照応要素として解析するものである。また、Kunoはゼロ代名詞を2種類に分類し(疑似ゼロ(pseudo-zero)、真ゼロ(real-zero))、それらについて別々に照応解析を行う方法を提案した⁵⁾。これは、全面的談話的省略(across-the-board discourse deletion)により省略された疑似ゼロ代名詞と、主題名詞句が談話場面に存在する名詞句が照応要素になり得る真ゼロ代名詞(英語における人称代名詞に類似)を、その照応要素の検出方法の点から分けて扱おうとするものである。

これらのゼロ代名詞の照応解析法は大きくみると、助詞のタイプや共感動詞の有無などの比較的表層的な情報を用いた上記のタイプと、プランやスクリプトの考え方を導入したタイプ⁷⁾に分けられる。機械翻訳への応用を考えたとき、前者は利用する情報が表層的なものに限られているため、解析精度の面で問題がある。また、後者は常識や世界モデルなどの膨大な知識を前提にシステムが構築されているため、幅広い分野のテキストを対象とした翻訳システムに適用するにはあらかじめ用意する知識の量が非現実的な規模となり、その実現は事実上困難である。また、ゼロ代名詞照応解析技術を機械翻訳システムに適用する際には、目的言語で訳出が必要となるゼロ代名詞を選択する必要も出てくる。よって、原言語と目的言語の言語的構造の違いを、照応解析が必要となるゼロ代名詞の認定アルゴリズムに反映させることが必須となる。

そこで、本論文では、用言意味属性に着目し、日本語談話中に現れるゼロ代名詞の照応解析を用言と用言の抽象的關係を用いて行う手法を提案する。本手法は、用言が持つ意味的属性をその動的特性と格に対する関係に着目して分類し、ゼロ代名詞を支配する用言の意味属性とその照応格要素を支配する用言の意味属性の2項関係によって、照応格要素を選択するものである。この方法は、機械翻訳システムのように対象分野を限定することができない文章でも、用言の意味的属性の対による翻訳対象分野とは独立したルールを記述すればよいので、少ない知識量を用意することに

よって、ゼロ代名詞の照応解析ができる。

2. 機械翻訳からみたゼロ代名詞

日本語談話には多数のゼロ代名詞が存在するが、照応解析が必要となるゼロ代名詞は、その解析結果の利用目的によって異なってくる。例えば、質問応答システムのように、ある文章を解析した上でユーザの質問に答えるような種類のタスクでは、質問対象の文章に対してさまざまな視点からの質問が来ることを想定しなければならないため、文章中のゼロ代名詞のほぼすべてについて解析が必要となる。それに対して、機械翻訳システムを考えた場合には、翻訳言語対の種類によって照応解析必要なゼロ代名詞は限定される。本章では、日英機械翻訳システムのタスクでゼロ代名詞を検出する際に考慮しなければならない点について述べる。日英機械翻訳システムでゼロ代名詞の検出の際に考慮すべき点は、その現象別に整理すると以下のよう

2.1 日本語表現と英語表現の発想の違い

機械翻訳でゼロ代名詞を検出する場合には、そのゼロ代名詞が、照応解析必要なものなのかを判断する必要がある。たとえば、

- (1) A社は2月1日、ハードディスク装置を発売する。

“Company A will put on sale the hard disk device from February 1.”

(ϕ subj ϕ obj) 月4百台生産する。

“ ϕ subj produce 400 unit of ϕ obj per month.”

という、文章の第2文では、用言「生産する」を中心とした構造となり、主語と目的語がゼロ代名詞化されている。しかし、より自然な表現で、英語に翻訳するためには、

- (2) 月産は4百台だ。

“Monthly production is 400 units.”

というような、名詞述語文(ダ文)の表現に書き替える必要がある。このように翻訳すると、「生産する」の主語と目的語のゼロ代名詞の照応解析をする必要がなくなる。このような表現の場合には、被翻訳文の構造解析時に、英語的な構造に変換した上で¹⁰⁾(J-J変換)、ゼロ代名詞を検索することによって、照応解析が必要かどうかの切り分けが可能となる。

2.2 日本語と英語の格パターンの違い

日本語では必須格となる格要素でも、英語に翻訳すると任意格となるような用言が存在する。たとえば、

(3) A (施設) で B (動物) を飼う*

“At A, B are being kept.”

という表現は、「飼う」の主語(ガ格)がゼロ代名詞化されているが、日本語で主語がなくても、“A raise B.”という表現で翻訳することが可能となる。このような表現で翻訳するためには、構文解析に用いる格パターンを、英語用言の訳語ごとに用意し、日本語解析時に英語の構文構造を決定しておくことが有効である。これによって、英語で必須格とならない要素は日本語では必須格と判断しないので、ゼロ代名詞の検出が正確に行える。

2.3 態による制約

日本語ではゼロ代名詞化されている要素でも、より自然な英語に翻訳をするためには、態を変換することによって、ゼロ代名詞の訳出が不要となる場合がある。たとえば、

- 日本文が受動態化されている文

この場合、英語表現としても受動態化することによって、照応解析に必要なゼロ代名詞を限定する。

- 日本語では受動態となる表現が、英語では能動態となるような用言からなる文

たとえば、

(4) A (*) が B (文書) に掲載される*

という表現は、「A を B に掲載する」の受け身表現であり、「掲載する」の主語(ガ格)がゼロ代名詞化されているが、英語では、日本語での主語がなくても、“A appear in B”という表現で翻訳するほうが日本語のニュアンスに近い訳が生成できる。このような場合には、構文解析に用いる格パターンを、英語の用法別に用意し、日本語解析時に、受け身の様相を伴った場合の必須格要素を英語の表現で必須格となる要素のみに制限することによって、照応解析が必要となるゼロ代名詞を限定する。

2.4 訳出構造による制約

- (5) A社はハードディスク装置を発売する。

(ϕ subj) ソフトを OS に組み込むことで接続台数を増やした。

という表現では、動詞「組み込む」、「増やす」のそれぞれの主語がゼロ代名詞化している。このうち、「組み込む」の文は、日本語では「こと」を修飾する埋め込み文の構造であるのに対し、英語に翻訳すると、「こと

で」が“by incorporating software into OS”という手段格となり、動名詞句としての表現になる。このように、日本語での埋め込み文を、動名詞句を伴う前置詞句化するように、日英で異なった文構造を生成する場合には、日本語解析時に英語的な構造に変換した上で、ゼロ代名詞を検出する。

日英機械翻訳システムでは、これらの条件を考慮した上でゼロ代名詞の認定を行う必要がある。

3. 新聞記事文でのゼロ代名詞の出現傾向

3.1 照応要素の位置関係

2章で示した条件を考慮した上で、新聞記事翻訳で問題となるゼロ代名詞とその照応要素の出現位置を調べた。一般に、新聞記事は表現を圧縮する場合が多く、用言活用語尾の圧縮による用言名詞化が頻繁に起こるため、助詞の情報や主題、共感動詞の種類のみではゼロ代名詞の照応要素を決定することができない場合が多い。たとえば、

- (6) NTT は 新型交換機 を 導入した。

“NTT will introduce a new model switchboard.”

(ϕ subj) 自己診断機能を 搭載、

“The new model switchboard is equipped with a self checking function and

(ϕ subj) 200 システム を 設置する 予定だ。

NTT is planning to install 200 systems.”

の例では、第1文の主語が主題化されており、第2文は、前半の文章の主語と後半の文章の主語がゼロ代名詞となっている。これらの二つのゼロ代名詞のうち、前者は前文の目的語である“新型交換機”が、後者は前文の主語である“NTT”が照応要素となる。このように、主題化された要素が存在し、文章中に他に主題化された要素がない場合でも、主題化された要素がゼロ代名詞の照応要素となるとは限らない。このような場合には、主題化されるか否かの情報だけではなく、もっと意味的な制約を導入することが必要となる。

3.2 新聞記事標本の分析

新聞記事文 30 記事のリード文、102 文に対して、日英機械翻訳システムで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。これによれば、第1文で助語“は”によって主題化された格要素が第2文以降で主語となるときゼロ代名詞化される例が 46 件 (49%) で最も多い。ま

* 括弧内は、用言による格への意味的制約を示し、*は意味的制約がないことを示す。

た、第一文に照応要素を持つゼロ代名詞の数が全体の80件(86%)にもものぼる。これは、新聞記事では第1文が記事全体の概要を示す情報を含むため、その格要素がゼロ代名詞の照応要素となる傾向の強いことを示している。その中でも、第2文以降のゼロ代名詞の照応要素が第1文にある例が71件(76%)もある点からも、第1文の重要性がわかる。また、ゼロ代名詞が主語であるが、その照応要素が主題や主語でない場合が15件(16%)もあることから、主題化された格要素から照応要素を選択する手法や、助詞の種類によって照応要素の優先順位を決定する手法だけでは、照応要素を選択するための情報としては不十分であることがわかる。さらに、この15件の例を詳細に見てみると、照応要素を支配するのは「結成する」、「発売する」、「開設する」、「開発する」というような、目的語の要素を新しく場面に導入するタイプの用言であり、ゼロ代名詞の用言は、「LANだ」のような名詞述語や、「所属する」のような属性を示す用言となっている。したがって、このようなタイプのゼロ代名詞の照応解析に対処するためには、用言が持つ属性を分類し、ゼロ代名詞の照応格要素を用言意味属性の関係から推測することが重要と判断できる。

4. 用言意味属性の分類

前章で述べたとおり、従来の手法では対処できない種類のゼロ代名詞照応も、意味情報を活用することによりかなり解決できると期待される。そこで本章では、実用的翻訳システムで取り扱える規模の知識を用いてゼロ代名詞の照応解析をするために、用言の持つ

意味属性を分類し、この属性間の関係から、ゼロ代名詞の照応要素を決定する。

本手法で採用する日本語の用言分類は、高松らの用言分類⁹⁾をもとにして、以下の二つの観点から拡張する。

用言分類基準

●用言が持つ動的特性

用言の持つ概念と談話場面に与える作用による分類

: 用言単体が発話されることにより、どのような種類の行為がなされ、その発話の結果どのような状況になるかという観点で分類。

Ex. 「持つ」→Possession

「開発する」→Production

「持つ」という用言は、Possessionの行為を行うことを文脈の状況の中に提示するのに対し、「開発する」という用言は、何かをProduceする行為を行うという事実を文脈の状況の中に提示する。

●用言の格に対する関係

用言の支配する格が用言に対してもつ役割による分類

: 格要素が支配される用言のもので発話される場合、その格要素が状況の中でどのような役割を果たすかの観点で分類。

Ex. 「完成する」

→SUBJが完成する: SUBJ be produced

「開発する」

→SUBJがOBJを 開発する

: SUBJ produce OBJ

表1 ゼロ代名詞とその照応要素の出現頻度

(調査対象文: 日経産業新聞, 情報・AI, リード文, 1988年2月分, 30記事(102文), 1記事あたり2-8文, 調査対象新聞記事中でゼロ代名詞を含む文は, 102文中56文)

Table 1 Frequency of appearance of zero pronouns and their referents.

(Source of sample sentences: Nikkei Sangyo Newspaper, Information and AI column, lead sentences during February, 1988. 30 articles (102 sentences) 2-8 sentences per article. Of the newspaper articles tested, the number of sentences with zero pronoun(s) contained was 56 out of 102.)

照応要素の出現場所 ゼロ代名詞の出現場所	第1文				第2文かそれ以降 ゼロ代名詞と同一文内				第2文かそれ以降 ゼロ代名詞と同一文外				文章中に無し	小計 [件]	
	は	が	を	他	は	が	を	他	は	が	を	他			
第1文	SUBJ	7	0	1	0	----	----	----	----	----	----	----	----	1	10
	OBJ	0	0	1	0	----	----	----	----	----	----	----	----	0	
	他	0	0	0	0	----	----	----	----	----	----	----	----	0	
第2文かそれ以降	SUBJ	46	4	15	0	6	0	0	3	0	0	0	0	3	83
	OBJ	0	0	6	0	0	0	0	0	0	0	0	0	0	
	他	0	0	0	0	0	0	0	0	0	0	0	0	0	
小計 [件]	80				9				0				4	93	

「完成する」、「開発する」ともに Production の行為を示す用言であるが、「完成する」は SUBJ を Produce するのに対し、「開発する」は SUBJ が OBJ を Produce することを示す。

この基準で分類した用言の概念体系を図 1 に示す。このような基準で分類された用言の属性を用言意味属性 (Verbal Semantic Attributes: VSA) と呼ぶ。これら用言意味属性は、用言を見出し語とする 2 種類のパターン辞書 (意味的結合価パターン変換辞書, 慣用表現変換辞書) に付与している。これらの辞書は、日本語用言 (慣用表現変換辞書においては、「油を売る」

のような慣用表現のコア部分) とその格要素への意味的制約からなる日本語単文パターンと、その日本語表現に対応する英語表現である英語パターンの対からなる。日本語と英語では統語構造が異なるので、このように 2 言語間で共通な語彙を示す表現を対で保持することにより、概念的な曖昧性の排除が可能となるため、詳細で正確な属性値の付与が可能となる。現在、約 15,000 対のパターン辞書の個々に対して VSA を付与している。個々のパターンの対には複数の VSA の付与が可能であるが、その約 9 割は 1 種類の VSA の付与で済んでいることから、概念的な曖昧性の解消

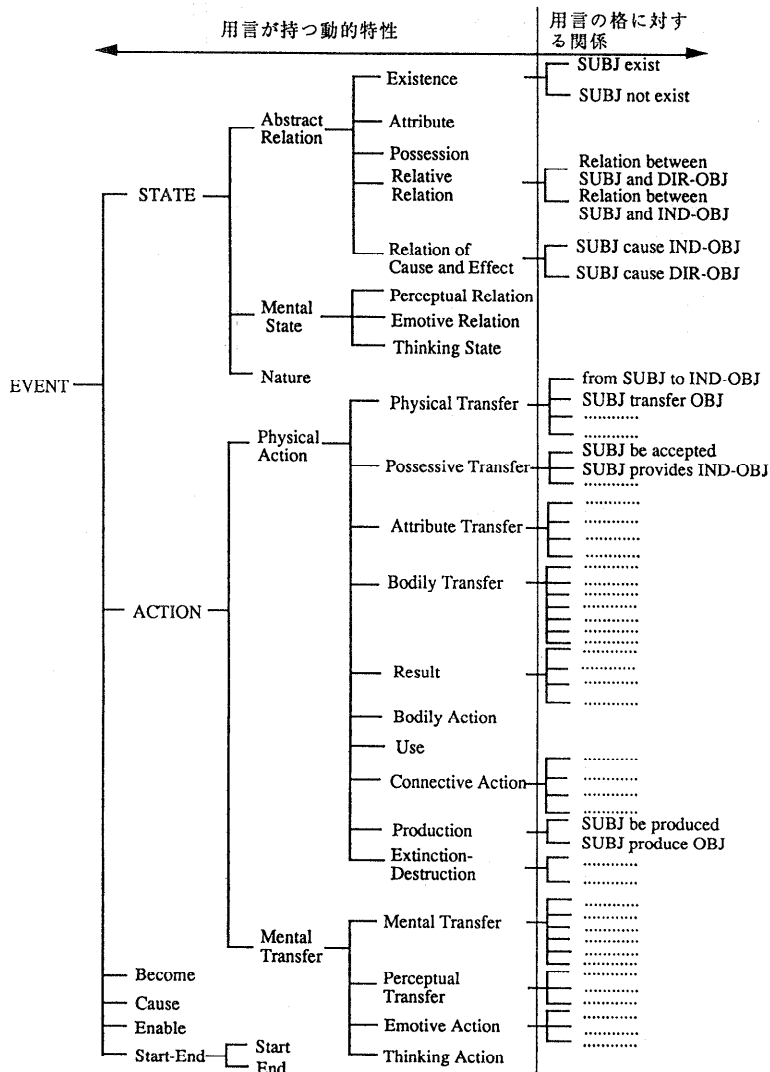


図 1 用言意味属性体系
Fig. 1 System of verbal semantic attributes.

が実現できているものと考えられる。

次に、ゼロ代名詞（省略格要素）がある文と、その照応格要素がある文（補完元）との意味的關係を、ゼロ代名詞の文がその照応格要素の文に対してどのような情報を新規に提供するかという観点で考える。この關係は、照応格要素を支配する用言とゼロ代名詞を支配する用言との用言意味属性の關係によって、表2のようにまとめることができる。この關係を用いれば、用言意味属性の2項關係によって用言關係の推測と、ゼロ代名詞の照応要素の決定が可能となる。この用言意味属性の關係決定のためのルールは、翻訳対象としている例文を用いて作成している。よって、現在のルールがすべての用言關係を網羅しているわけではない。しかし、そのルールは、その翻訳対象となる分野で出現する普遍的な用言關係を規定したものであり、ルール自体の分野依存性は少ないと言える。翻訳対象の文を変更した場合には、不足となるルールを追加する必要があるが、すでに作成されているルール群は普遍的なものであり、ルール自体も用言意味属性の2項關係という極めて単純なルール構成となっているため、スクリプトを新規に作成する等のその他の手法に比べるとルール作成のための負荷はかなり軽減できる。

3章で述べたとおり、新聞記事リード文は第1文が記事全体の内容を概説する文章構造を持っていることが多く、第1文内の単位文が主題化されていることが多い。ここではこのような単位文を主題単位文と呼び、この主題単位文の用言意味属性によって、以降の文のゼロ代名詞の照応要素を選択していく。また、用言關係を決定するルール群には優先順位が付与されており、個々のルールはこの優先順位順に適用される。

上記に示すように、用言意味属性の分類体系の属性値で記述したゼロ代名詞の照応要素の決定ルールを用いることによって、汎用的な少ない知識量で照応解析処理系が実現できるので、実用的に分野の限定が困難な機械翻訳への応用が可能となる。

5. 照応解析方式

5.1 アルゴリズム

上記の用言意味属性を用いた、ゼロ代名詞の照応解析を行うためのシステム構成を図2に示す。解析対象の日本語はすでに形

表2 用言意味属性による照応關係決定ルール
Table 2 Rules for determining resolution elements by verbal semantic attributes.

ゼロ代名詞の条件	照応要素の条件	用言關係	決定された照応要素
V SA	格	V SA	
POSS	SUBJ	POSS-TRANS2 & START	OBJ
THINK-ACT	SUBJ	POSS-TRANS2 & START	SUBJ
.....

態素解析、構文解析、意味解析が行われているとし、その結果が文脈解析の対象入力となる。文脈解析では、次のような流れでゼロ代名詞の照応解析が行われる。

(ステップ1)

ゼロ代名詞を検出する。

もし存在すれば、同一文内にその照応格要素が存在しないか調査する。

存在し、照応解析が終了すればステップ4へ。

同一文内の照応格要素の検出は

- 1) 接続詞の種類による、ゼロ代名詞の照応解析
- 2) 用言意味属性による、照応解析

の2種類を採用している。

1)は二つの単文から構成される文（複文や重文）において、助詞の種類と、接続語の種類によって、照応要素が構文的に決定するような制約を用いたものである⁹⁾。接続語の種類による照応要素決定ルールの一部を表3に示す。

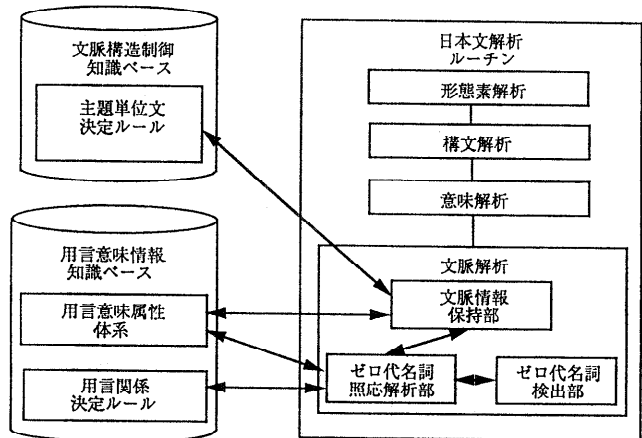


図2 本システムの構成
Fig. 2 Structure of this system.

表 3 接続語によるゼロ代名詞とその照応要素との条件
Table 3 Constraints to zero pronouns and their referent with connecting words.

接続語の例	ゼロ代名詞の条件	照応要素との関係*
から, し, ば	ハ格	従文→主文
ため	ハ格	従文←→主文
まま	ハ格, ガ格	従文→主文
たり, て	ハ格, ガ格	従文←→主文
と	ハ格, ヲ格	従文→主文
つつ, ながら**	ハ格, ガ格, ヲ格	従文←→主文

* この矢印は、照応要素の含む文から、これと照応可能なゼロ代名詞を含む文への方向を示す。

** “つつ”, “ながら” の場合、ヲ格はその接続関係が「逆接確定」の場合のみ補完対象となる。

2) は同一文内の、接続語によって、照応要素が決定できない場合、(例えば、3種類以上の単文が1文中に存在するような場合) 用言意味属性を用いて、照応解析を行うものである。

(ステップ2)

同一文内に存在しなければ、文脈情報保持部に保持されている主題単位文の格要素から、照応格要素候補を選択する。選択基準はゼロ代名詞が支配する用言の用言意味属性と主題単位文の用言意味属性の関係で、表2の用言関係決定ルールとする。

用言による制約を満たせば、照応関係が成立し、ステップ4へ。

(ステップ3)

照応格要素が検出できない場合、補完不可能として、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。

(ステップ4)

文脈構造制御知識ベースから、解析対象分野の文章構造に依存して決定した主題単位文決定ルール*を用いて、主題単位文を選択し、文脈情報保持部に保存。

次の文章へ。

5.2 例

(6)の例文を用いて本手法によるゼロ代名詞の照応解析の例を(7)に示す。

(7) [第1文] NTT は 新型交換機 を 導入する。

“NTT will introduce a new model switchboard.”

主題単位文:

(introduce (VSA (POSS-TRANS 2 & START))
(SUBJ “NTT”)

(OBJ “new model switchboard”)

[第2文] ϕ subj 自己診断機能を搭載、

“ ϕ subj is equip with a self checking function and

<解析結果> (equip (VSA (POSS))

(SUBJ ϕ SUBJ)

(OBJ “a self checking function”)

ϕ SUBJ=“new model switchboard”

ϕ subj 200 システム を 設置する 予定だ。

ϕ subj is planning to install 200 systems.”

<解析結果> (be-planning-to (VSA (THINK-ACT))

(SUBJ ϕ SUBJ) (OBJ...))

ϕ SUBJ=“NTT”

主題単位文:

(introduce (VSA (POSS-TRANS 2 & START))
(SUBJ “NTT”)

(OBJ “new model switchboard”)

第1文の解析後、解析結果より主題単位文が抽出される。この例文では、第1文が単位文で構成されるので、その文章の解析結果が主題単位文として文脈情報保持部に保持される。次に、第2文の解析結果より、「搭載する」と「予定だ」の主語がゼロ代名詞化されていることが分かる。同一文内にはそれらの照応格要素が存在しないので、主題単位文中の格要素が照応格要素候補となる。「搭載する」と「予定だ」のVSAはそれぞれ“POSS”, “THINK-ACT”であり、主題単位文の用言のVSAは“POSS-TRANS 2 & START”であるので、表2のルールにより用言意味関係は「詳細説明」「方針決定」が成立し、主題単位文の目的語、主語がそれぞれの照応要素となる。

6. 機械翻訳システムへのインプリメン

ト^{11), 12)}

本論文で提案した手法をインプリメントした日英翻訳システム ALT-J/E の処理の概要を説明する(図3)。最初に、入力日本語は形態素解析の後、文要素間の係り受け関係の解析を行う。用言の間の係り受け関係の認定の結果に基づいて単位文を抽出し、さらにそ

* 新聞記事の場合には記事の第1文が主題単位文となる。第1文が複数の単位文からなる場合は、接続語の種類によって、主題単位文に優先順位を付与する。具体的には、複文の場合は主文を優先させる等のルールを用いる。

の中から単文を抽出する^{*}。単文は述部より様相、時制などの主体的表現情報を抽出し、これらを抜き取った平文に変換する。この平文は図4に示すように用言を見出し語とする2種類のパターン辞書(慣用表現変換辞書, 意味的結合価パターン変換辞書)と順に照合され、適切なパターンのない場合には汎用パターン変換規則の適用を受ける。これにより、日英変換で使用される文型パターンを決定する。なお、2章の(3)や(4)のような場合には、処理上はゼロ代名詞とは認定されない。この時点で多義が残る場合には平文レベルでのパターンに対する解釈の結果に基づいて最終的に一つの解釈を決定する。また、2章の(1)や(5)のように、日本語の構造と英語の構造が大幅に異なる場合には、解析結果の日本語構造を英語の表現に近い日本語構造に変換することにより、英語翻訳時に不必要なゼロ代名詞の照応解析をさける。上記の解析の結果、原文中にゼロ代名詞が存在する場合には、5章で示したとおり、文脈情報により格要素補完を行う。この時点で、単位文生成で使用される文型パターンは決まっているため、あとは、それを適用し、英語の骨格的表現である素文を生成し、様相、時制、接続などの情報を付加する。その際には、照応解析後のゼロ代名詞の訳出方法を冗長な英文にならないように省略や代名詞化、定冠詞化等を施し、注意して決定する¹³⁾。

7. 評 価

2章で用いた新聞記事リード文 30 記事 102 文を処理対象文に、本手法を適用した場合のゼロ代名詞の出現場所とその解析成功率を表4に示す。この102文は、照応解析のためのルールを整備しており(実際に追加修正したルールは、用言意味属性を用いたルール 116 ルール中の文内ルール 3 ルール、文間ルール 6 ルールのみで、接続詞の種類を用いたルール (23 ルール) は、この処理対象文のために特に整備していない)、この結果は本方式の技術的限界を示すウインドウテストの結果と言える。

2章で示したとおり、本処理対象文において、方式上、本手法で適用対象となるゼロ代名詞(テキスト文

^{*} 単位文は文を述部中心の木構造に展開した際、最上位のレベルでの述語が単一の文であり、埋め込み文等を含むこともある。これに対して単文は単位文を単一の用言しか持たないレベルに分解したものを言う。

- (1) 形態素解析……………分かち書き, 単語文法カテゴリの決定
- (2) 係り受け解析……………文要素間の関係の決定
- (3) J/J 変換……………日本語内での表現変換
- (4) 単文抽出……………係り受け解析結果から各用言の支配範囲を決定
- (5) 単文解析
 - (6) 述部解析……………様相等を抽出し, 平文化する
 - (7) 名詞句解析……………名詞句, 複合語の意味構造の決定
- (8) 埋め込み文解析……………連体修飾等の意味構造を決める
- (9) 平文英語変換……………パターン対辞書で客体的表現を変換
- (10) 接続解析……………用言間の関係を決める
- (11) 最適結果選択……………文全体の多義を絞り込む
- (12) 省略格要素補完……………省略された格要素を文脈情報により補完する
- (13) 補完要素変換……………補完された要素の変換方法を決定する
- (14) 単位文生成
 - (15) 基本構造生成……………英文全体の構造を決定する
 - (16) 副詞句生成……………様相, 時制, 動詞等から副詞句訳決定
 - (17) 名詞句生成……………句構造と複合語構造の変換, 埋め込み文組込み
- (18) 接続構造生成……………接続属性, 主語の有無等に応じて単位文を接続
- (19) 様相・時制構造生成……………助動詞, 不定詞の挿入, 語型/構文変形など
- (20) 英文調整……………縮約, 外置変形, 冠詞付与, その他

図3 日英翻訳システム ALT-J/E の処理概要

Fig. 3 Process outline of Japanese-English machine translation system, ALT-J/E.

[慣用表現の例]

- (1) 慣用句パターンの例
N1 (主体) は背が高い => N1 be tall
- (2) 機能動詞結合の例
N1 (主体) が N2 (主体) の非難を浴びる。
(-> N1 が N2 から非難される) 日本語内での
(-> N2 が N1 を非難する (+受動)) 変換
(=> N2 claim N1 (+受動))……………日英変換パターン適用
=> N1 be claimed by N2 ………………英語変形

[意味的結合価パターン例]

- N1 (主体) が N2 (文化/人間活動) を暗記する。
=> N1 learn N2 by heart
N8 (施設) で N2 (動物) を飼う。=> N8 raise N2
N1 (主体) が N2 (食料) を食べる。=> N1 eat N2
(<かわとり) (chicken)

- (1) 鳥……………hen
- (2) 食料……………chicken

図4 日英変換パターン辞書の適用例

Fig. 4 Example of application of Japanese-English conversion pattern dictionaries.

中にその照応要素が現れる)の割合は全体の約96%であり、本手法の適用範囲の広いことがわかる。その適用対象のゼロ代名詞のうち照応解析に成功した割合は約99%、適用対象外のゼロ代名詞を含んだ割合でも約95%と極めて高い値を示した。

本手法で照応解析に失敗した例を以下に示す。下記のとおり、これら失敗例のうちの4件は文章中に現れない要素と照応関係にあるゼロ代名詞であった。たとえば、(9)においては、オフコン(オフィスコンピュータ)を用いたシステムの販売開始の文脈の中で用いら

表 4 本手法で照応解析に成功したゼロ代名詞の頻度 (ウインドウテスト)
(調査対象文: 表 1 と同様. 照応解析ルール整備済み)

Table 4 The frequency of successful resolution of zero pronouns by this method. (window test)

(Source of sample sentences: same as Table 1. Rules for anaphora resolution are already tuned.)

照応要素の出現場所 ゼロ代名詞の出現場所		第 1 文				第 2 文か それ以降 ゼロ代名詞と同一文内				第 2 文か それ以降 ゼロ代名詞と同一文外				文章中 に 無し	小計 [件]
		は	が	を	他	は	が	を	他	は	が	を	他		
第 1 文	SUBJ	7 / 7	0	1 / 1	0	--	--	--	--	--	--	--	0 / 1	9 / 10 90%	
	OBJ.	0	0	1 / 1	0	--	--	--	--	--	--	--	0		
	ETC.	0	0	0	0	--	--	--	--	--	--	--	0		
第 2 文 か それ以降	SUBJ	46 / 46	4 / 4	15 / 15	0	6 / 6	0	0	2 / 3	0	0	0	0	0 / 3	79 / 83 95%
	OBJ	0	0	6 / 6	0	0	0	0	0	0	0	0	0		
	ETC.	0	0	0	0	0	0	0	0	0	0	0	0		
小計[件]		80 / 80 100%				8 / 9 89%				0				0 / 4 0%	88 / 93 95%

* 表中の分母はゼロ代名詞の出現件数, 分子は照応解析に成功したゼロ代名詞の件数を示す。

れ, 「オフコンになって」の「なる」の主語がゼロ代名詞化している。この文章は, このシステムで用いられているコンピュータの主流がオフコンであるという文章であるため, このゼロ代名詞の照応要素は文章中には現れない「主流の機種」となる。また(10)においては, ある時期の電線の出荷量の統計結果を示す記事の中で, 「好調を続けている」の主語がゼロ代名詞化しているが, この主語も文章中には現れず, 「電線業界」がその照応要素となる。これらを取り扱うためには, 世界知識等を用い深い推論を行うことによって, 照応要素を推測するように機能拡張する必要がある。また残りの 1 件は, 照応要素は文章中に現れるが格要素ではない場合である。具体的には, (11)に示すとおり, 「~のもの」の主語がゼロ代名詞化しており, その照応要素が前文の目的語(ヲ格)「~システムの販売」の「~システム」となる場合に照応解析に失敗した。この場合, J-J 変換により, 主題単位文の“~システムの”の文章構造を“~は~システムを販売し始める。”という文章のように格要素化することで, 本手法でも正しく解決することが可能となる。

● 補完処理失敗例 (ウインドウテスト) 5 件

- (a) 文章中に現れない要素がゼロ代名詞の照応要素となる。 4 件

例: (9)“(主流の機種が)* オフコンになって, ~”
(10) “A社のまとめた電線統計によると, ~と (電線業界が)* 好調を続けている。”

(b) 名詞に「の」で修飾する名詞がゼロ代名詞の照応要素となる。 1 件

例: (11) “~は~システムの販売を始める。”
“(~システムは)* ~のもの。”

次に, 前記の 102 文のために整備したルールを用いて, この 102 文とは異なる無作為に抽出した新聞記事 98 文について, 照応解析を行った結果を表 5 に示す。本試験文において, 文章中に照応要素が出現するゼロ代名詞の割合は約 88% であった。ルールをこれらの文に整備せずにブラインドテストの状態で本手法で照応解析を行った結果, 照応要素が文章中に現れるゼロ代名詞のうち約 83% が照応解析に成功し, 照応要素が文章中に現れないゼロ代名詞を含んだ場合でも約 73% の成功率を達成した。

なお, 照応要素が文章中に存在する 59 件中で照応解析に失敗した 10 件のうち 6 件は, 照応解析ルールを追加することによって, また 1 件は, 構文解析に用

* 括弧で囲まれた格要素は, それがゼロ代名詞化しており, それがゼロ代名詞と照応する要素であることを示す。

表 5 本手法で照応解析に成功したゼロ代名詞の頻度 (ブラインドテスト)
 (調査対象文: 日経産業新聞, 情報・AI 欄, リード文, 1988年2月分, 33記事 (98文), 1記事あたり1-5文, 調査対象新聞記事中でゼロ代名詞を含む文は, 98文中45文. 照応解析ルール未整備.)

Table 5 The frequency of successful resolution of zero pronouns by this method. (blind test)

(Source of sample sentences: Nikkei Sangyo Newspaper, Information and AI column, lead sentences during February, 1988. 33 articles (98 sentences) 1-5 sentences per article. Of the newspaper articles tested, the number of sentences with zero pronoun(s) contained was 45 out of 98. Rules for anaphora resolution aren't tuned for these sentences.)

照応要素の出現場所 ゼロ代名詞の出現場所		第1文				第2文かそれ以降 ゼロ代名詞と同一文内				第2文かそれ以降 ゼロ代名詞と同一文外				文章中に 無し	小計 [件]
		は	が	を	他	は	が	を	他	は	が	を	他		
第1文	SUBJ	9 / 9	0	0	0	--	--	--	--	--	--	--	--	0	11 / 11 100%
	OBJ.	0	0	2 / 2	0	--	--	--	--	--	--	--	--	0	
	ETC.	0	0	0	0	--	--	--	--	--	--	--	--	0	
第2文 か それ以降	SUBJ	32 / 35	0	3 / 5	0 / 1	2 / 2	0	0 / 2	0 / 1	0	0	0	0	0 / 7	38 / 56 69%
	OBJ	0	0	1 / 1	0	0	0 / 1	0	0	0	0	0	0	0 / 1	
	ETC.	0	0	0	0	0	0	0	0	0	0	0	0	0	
小計[件]		47 / 53 89%				2 / 6 33%				0				0 / 8 0%	49 / 67 73%

いた意味的結合価パターンの変換パターン不足が原因であるので変換パターンの追加にすることによって解決する。その結果, 正しく照応解析ができないものは3件にとどまり, 容易に95%の品質を達成できる。

このように, 特定の標本を対象に準備したルールをその他の標本に適用した場合でも, 文章中に照応要素が出現するゼロ代名詞の8割強にはそのまま適用できることから, ルールの汎用性が高いことがわかる。また, 若干のルール追加によって95%の品質を得ることができる点からそのルール整備のための労力は少なく済むと言える。

8. ま と め

本論文では, 日本語談話中に現れるゼロ代名詞に対して用言意味属性を用いた照応解析の方法を提案した。本手法は, 用言が持つ意味的属性をその種類により分類し, その属性値の2項関係でルールを記述したため, 比較的少ない知識量でゼロ代名詞の照応を解析することが可能となった。本手法を機械翻訳システムALT-J/Eに組み込み, 新聞記事を対象とする2種類

の性能評価を行った結果, 日英翻訳で照応解析が必要となるゼロ代名詞のうち, 照応解析ルールを整備した標本実験では全体の95%が, また, この実験で使用したルールをそのまま適用した別の標本実験では全体の73%がその照応要素を正しく解析できることがわかり, 本手法の有効性が実証できた。

なお, 本手法の文脈処理機能の応用としては, 記事全体の文章の構造解析による要約文の生成や, 第1文の文章中の用言の意味属性に着目した記事の内容分類などが考えられる。

今回の報告では, 処理対象文を新聞記事リード文と比較的短い文章を対象としたが, 今後, 話の変わり目や, 談話構造がより複雑な文章についても検討が必要である。また, 今後ルール数を増加させた場合には, ルール適用の曖昧性が生じることが考えられるため, 翻訳対象分野に応じた適用ルールの優先度の制御などの検討も必要である。さらに, 照応要素が文章中に現れないゼロ代名詞の照応要素の推定についても検討が必要である。

謝辞 本研究を進めるにあたりご指導いただいた白

井論主任研究員に感謝いたします。また、日頃熱心に討論していただく NTT 情報通信網研究所知識処理研究部池原研究 G の皆様に感謝します。

参 考 文 献

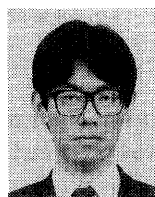
- 1) 久野 暉：談話の文法，大修館書店 (1978)。
- 2) Grosz, B. J. et al.: Providing a Unified Account of Definite Noun Phrases in Discourse, *Proceedings of 21st Annual Meeting of the ACL*, pp. 44-50 (1983).
- 3) Kameyama, M.: A Property-Sharing Constraint in Centering, *Proceedings of 24th Annual Meeting of the ACL*, pp. 200-206 (1986).
- 4) Yoshimoto, K.: Identifying Zero Pronouns in Japanese Dialogue, *Proceedings of 12th International Conference on Computational Linguistics*, pp. 779-784 (1988).
- 5) Kuno, S.: Identification of Zero-Pronominal Reference in Japanese, *Proceedings of ATR Symposium on Basic Research for Telephone Interpretation*, pp. 5-2-1-5-2-19 (1989).
- 6) Walker, M. et al.: Centering in Japanese Discourse, *Proceedings of 13th International Conference on Computational Linguistics*, pp. 1-6 (1990).
- 7) 堂坂浩二ほか：日本語省略理解のための手法とその実現法，情報処理学会自然言語処理研究会資料，47-3 (1985)。
- 8) 高松 忍，西田富士男：動詞パターンと格構造に基づく英日機械翻訳，電子通信学会論文誌，Vol. J 69-D, No. 9, pp. 815-822 (1981)。
- 9) 南不二男：現代日本語の構造，大修館書店 (1974)。
- 10) 白井 論：日本文自動書き替えによる構文多義の解消，第 41 回情報処理学会全国大会論文集，4 S-6 (1990)。
- 11) 池原 悟ほか：言語における話者の認識と多段翻訳方式，情報処理学会論文誌，Vol. 28, No. 12, pp. 1269-1279 (1987)。
- 12) Ikehara, S. et al.: Toward an MT System

without Pre-Editing—Effects of New Methods in ALT-J/E, *Proceedings of MT Summit-III*, pp. 101-106 (1991).

- 13) 中岩浩巳：日英翻訳システムにおける日本語省略補完格要素の訳出方法，1990年人工知能学会全国大会論文集，pp. 323-326 (1990)。
- 14) 中岩浩巳：接続関係情報・用言意味情報を用いた日本語省略格要素補完，1989年人工知能学会全国大会論文集，pp. 371-374 (1989)。
- 15) Nakaiwa, H. and Ikehara, S.: Zero Pronoun Resolution in Japanese to English Machine Translation System by using Verbal Semantic Attributes, *Proceedings of the 3rd Conference on Applied Natural Language Processing*, pp. 201-208 (1992).

(平成 4 年 10 月 26 日受付)

(平成 5 年 6 月 17 日採録)



中岩 浩巳 (正会員)

1963 年生。1985 年法政大学工学部電気工学科卒業。1987 年名古屋大学大学院工学研究科電気系専攻博士前期課程修了。同年日本電信電話(株)入社。以来、日英機械翻訳技術の研究に従事。現在、NTT 情報通信網研究所知識処理研究部研究主任。人工知能学会、ACL 各会員。



池原 悟 (正会員)

1944 年生。1967 年大阪大学基礎工学部電気工学科卒業。1969 年同大学大学院修士課程修了。同年日本電信電話公社に入社。現在、NTT 情報通信網研究所知識処理研究部勤務。この間、数式処理、トラヒック理論、自然言語処理の研究に従事。工学博士。1983 年情報処理学会論文賞受賞。電子情報通信学会、人工知能学会各会員。