自然言語処理 123－5
（1998．1．19）

日英対訳コーパス中の ゼロ代名詞とその指示対象の自動認定

中岩浩巳
NTTコミュニケーション科学研究所

従来からのゼロ代名詞の照応解析技術では照応解析規則を人手で作成する必要があり，規則の蓄積に時間的・人的コストがかかっていた．本論文では，このゼロ代名詞照応解析規則の自動抽出を目指して，日本語と英語ではゼロ代名詞の出現傾向の大きく異なるというに着目した日本語文中のゼロ代名詞とその英語文中の指示対象を自動抽出する手法について提案する．本手法では，まず日英対訳文対中から対訳関係にある日英表現対を抽出した後，10種類の規則を用いて日本語文中のゼロ代名詞と英語文中の指示対象の英訳表現を抽出する．日英機械翻訳評価用例文集中の554種類のゼロ代名詞に対して提案手法を適用したところ，91.5%のゼロ代名詞とその指示対象が正しく認定でき，本手法の有効性が実証できた．

Automatic Identification of Zero Pronouns and their Antecedents within Aligned Sentence Pairs

Hiromi Nakaiwa
NTT Communication Science Laboratories

This paper proposes a method to identify zero pronouns within a Japanese sentence and their antecedent equivalents within the corresponding English sentence from aligned sentence pairs. The method focuses on the characteristics of Japanese and English, in two languages from diffierent families and in which distribution of zero pronouns is very different. In this method, the pairs of Japanese word/phrase and their English equivalent word/phrase are identified. Next, zero pronouns within a Japanese sentence and their antecedents within the English sentence are identified by using the characteristics of anaphoric and deictic expressions in English. According to my evaluation, for 554 zero pronouns in a sentence set for the evaluation of Japanese-to-English machine translation systems, 91.5% of the pairs of zero pronouns in the Japanese sentences and their antecedents in the English translations were automatically identified correctly.

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．

日本語ゼロ代名詞の照応解析に関しては，従来から様々な手法が提案されてきているが[1][2][3][4]，翻訳対象分野を限定しない機械翻訳システムに応用することを考えると，解析精度の点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点で問題があり，実現は困難である．これらの問題に対しては，照応解析条件として，用言の意味属性[5][6]，様相表現，接続表現を用い，これらを表現の持つ意味に応じて分類し，その代表的属性値に応じて照応要

素を決定することによりこれらの問題を考慮にいれた，機械翻訳に適した照応解析手法が提案されている[7][8][9].

しかし，これら従来から提案されている手法では，基本的に人間が照応解析のための規則を作成する必要がある．よって，網羅的な照応解析規則を作成するためには，かなりの専門知識と労力が必要となる．さらに，解析対象となる分野に応じて，異なった照応要素を認定する必要があるゼロ代名詞が存在するので，分野に依存した照応解析規則を作成する必要がでてくる．しかし，分野毎にこの規則を作成することは，このための労力や時間を考慮すると，実際的には実現不可能である．よって，このゼロ代名詞の照応解析規則を効率的に獲得する手法の実現が望まれている．

自然言語処理システムにおける解析規則の効果的な獲得のためには，従来から，既存のコーパスを用いて，コーパス中に現れる言語現象を分析し，分析した結果を基に解析規則を抽出する手法が提案されている．ゼロ代名詞照応解析規則の自動抽出に関してもいくつかの手法の提案がされてきており，それらは多くは基本的に解析対象となる言語のコーパスのみを用いている[10][11]．しかし，解析対象言語のコーパスのみを使用する場合，その言語ではほぼ常にゼロ化される要素を補完するための規則を抽出することは困難である．また，解析対象となるタイプのゼロ代名詞の指示対象を決定するための情報を含む言語現象が，その解析対象文以外の文中に現れなければ，有効な解析規則を抽出できない．

このような問題を考慮にいれると，ゼロ代名詞の解析規則の抽出に用いるコーパスとしては，解析対象の言語のみからなるコーパスではなく，解析対象の言語と他の言語の対訳コーパスを利用することが有望であると期待できる．特に，日本語と英語のように言語族が異なる場合には，省略現象が現れる傾向が異なるため，ある言語の文ではゼロ化されている要素が，その文と対訳関係にある別の言語の文では明記される場合が多々有り，その利用が有望である．

対訳関係にある文の集合である対訳コーパスから各種解析規則を抽出するためには，それら集合から，対訳関係にある文対を抽出する技術，対訳関係にある文対から対訳関係にある単語・表現対を抽出する技術が重要となり，従来から様々な手法が提案されてきている．ゼロ代名詞照応解析規則の抽出という観点で考えると，対訳関係にある一方の文からゼロ代名詞を抽出し，他方の文からそのゼロ代名詞の指示対象を抽出する技術が必要となる．これに関連した技術としては，最近，対訳コーパス中の段落や図，表など文章単位での省略箇所を抽出する手法が提案されているが[12]，文中でゼロ化された箇所とそこに補うべき要素の抽出に関する手法の提案はされていない．

本稿では，このような目的を達成するための第１段として，１文対１文で対訳関係にある日英の対訳文からなる日英対訳コーパスから，日本語文中のゼロ代名詞と，英語文中のそこに補うべき照応要素を抽出する手法を提案する．

２．ゼロ代名詞照応解析規則の自動抽出

本章では，日英対訳コーパスからゼロ代名詞照応解析規則を自動抽出するシステムの全体構成について述べる[1]（図１）．図に示すとおり，日英対訳コーパス中の日本語文とその英語文は，日英別々に日本語解析システム及び英語解析システムで解析される．次に，日英の解析構造を参考にして，その日英対訳文中の対訳関係にある日英表現対を抽出する．次に，日本語ゼロ代名詞およびそれに補う英語指示対象を抽出する．そして，英語補完要素をもとに，日英対訳辞書等を用いて，日本語指示対象を抽出する．以上の結果をもとに，

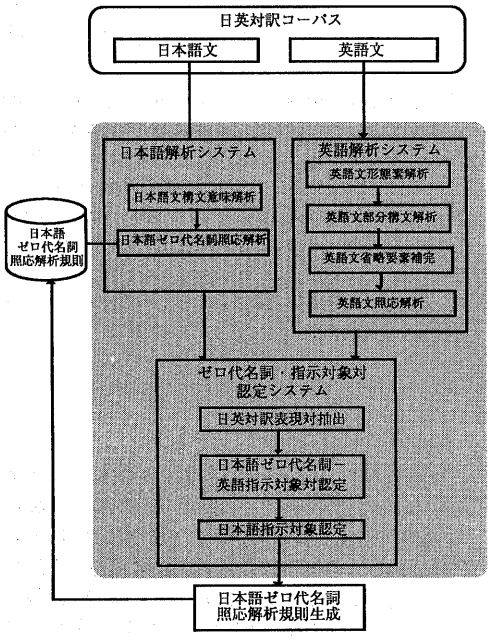


図１　ゼロ代名詞照応解析規則自動抽出の構成

[1] 処理の詳細に関しては，文献[13][14]を参照のこと．

日本語解析結果を参考に，省略要素補完規則を作成する．その後は，その新規作成された省略要素補完規則を用いて，同じ対訳コーパスを対象に，新規作成された省略要素補完規則の有効性を検証しつつ，再度この学習過程を繰り返す．

3．対訳文中のゼロ代名詞と指示対象

よく知られているように，日本語では，聞き手もしくは読み手が文脈や常識から容易に推測できる場合には，主語や目的語等の格要素は省略される場合がほとんどである．それに対し，英語では，明示される場合がほとんどである．例えば，

(1) (ϕ-が) 本を読みたい
"I want to read a book."

という表現では，希望の様相表現「たい」を伴うため，読み手もしくは聞き手は，特別な文脈がなければ，「本を読みたい」のは，話し手か書き手であることが容易に推測できるため，動詞「読む」のガ格である「私」が省略されている．しかし，このような場合でも，英語では，動詞"read"の主語として，"I"が明示される．

本節では，対訳関係にある日本語文と英語文を用いて，日本語文中のゼロ代名詞と，英語文中で明示する必要のあるゼロ代名詞の指示対象の英訳表現の傾向について考察する．

3．1 調査対象文

本調査では，日英機械翻訳システム評価用例文3718文の日英対訳例文集[15]を用いて，日本語文中のゼロ代名詞と英語文中のゼロ代名詞に補うべき要素の英訳表現を考察した．この評価用例文には，文内照応ゼロ代名詞と文章外照応ゼロ代名詞が多く出現するため，日本語ゼロ代名詞とその指示対象の英訳表現を検討するのに適している．また，個々の例文は自然な日本語文であり，様々な日本語表現の翻訳性能の評価の目的で，広範囲な表現が含まれているため，様々なタイプのゼロ代名詞が英語表現にいかに翻訳されているかの情報を得ることが出来る．なお，本例文の平均文長は日本語文で13.39文字，英語文で9.22単語である．

3．2 出現傾向

上記の試験文に対してゼロ代名詞とその指示対象の出現傾向を調査した結果を表1に示す．指示対象は出現場所から見て，同一文内に存在する場合と，同一文内に存在しない場合も分かれる．このうち同一文内に存在しない場合におけるゼロ代名詞は，英語への訳し方から見て，以下のように分類することが出来る．

- 受動態に変換しゼロ代名詞となる要素を訳出させない
- 指示対象が筆者・話者"I"かそのグループ"we"
- 指示対象が読者・聴者"you"
- 指示対象がだれか特定できない人間
- 指示対象として"it"に訳される
- 指示対象として"this"に訳される
- その他の特定の要素が指示対象になる

調査結果によれば，全ゼロ代名詞554件に対して150件(27%)の指示対象が同一文内に現れ，404件(73%)の指示対象は文中に現れない．この404件のうち，受動態に翻訳することによって指示対象を認定せずに生成されたゼロ代名詞は157件(28%)にとどまる．よって残りの247件(45%)は，指示対象が日本語文中には存在しないが英語文中には存在する．この結果から，対訳例文集は，日本語文中のゼロ代名詞と英語文中の指示対象を認定して，日本語ゼロ代名詞とその指示対象の対を自動認定することに有効であることが分か

表1 ゼロ代名詞とその指示対象の出現頻度

ゼロ代名詞化された格要素	指示対象の出現場所												小計[件]
	同一文内					同一文内になし							
	は	が	を	に	他	受身	Iかwe	you	人間	it	this	他	
は	2	0	0	0	0	3	1	2	0	1	0	0	9
が	113	12	1	1	8	152	88	25	31	57	3	0	491
を	4	0	6	1	0	0	0	0	2	19	1	0	33
に	1	0	0	0	0	2	3	8	0	1	0	2	17
の	0	0	1	0	0	0	1	0	0	2	0	0	4
小計[件]	150					404							554

る.

表1の日本語ゼロ代名詞の英訳表現を詳細に分析してみると,指示対象のタイプに応じてその英訳表現のスタイルも異なってくる.その特徴は次のようにまとめることが出来る.

- 指示対象が文中に存在しない直示的(deicitc)ゼロ代名詞(247件) : "I"や"you"のような人称代名詞が代名詞,不定詞"one"で翻訳.
- 指示対象が文中に存在する照応的(anaphoric)ゼロ代名詞(150 件;文内照応) : 人称代名詞に加え,"that"等の指示詞,"the company"等の定冠詞を伴う定表現,照応的"one"で翻訳.

よって,英語文中のこのような表現に着目することによってゼロ代名詞の指示対象候補を効果的に抽出できることが期待される.

4. 自動認定手法

3章での議論をもとに,本節では,1対1で対訳関係にある文対からなる日英対訳コーパスから,日本語文中のゼロ代名詞と,そのゼロ代名詞に補完すべき英語文中の指示対象を自動認定する手法について提案する.図1の網かけされた部分が本手法の全体構成である.本処理は次の3種類の部分からなる;(a) 日本語文の解析,(b) 英語文の解析,(c) 日本語文中のゼロ代名詞と英語文中の指示対象の認定.個々の処理について次節以降で詳細に述べる.

4. 1 日本語文解析処理

日本語文は,NTTが研究開発中の日英機械翻訳システムALT-J/E[16][17]の形態素解析,構文意味解析処理系によって解析される.日本語文解析は以下の手順で行われる.

[Step 1] 日本語文の形態素解析

日本語文は単語分割され品詞情報が付与される

[Step 2] 日本語文の構文意味解析

日本語文中の各単語とその品詞情報を用いて,構文意味構造が生成される.ALT-J/E では,日本語の構文意味構造は,直接翻訳する英文の構造と対応が取られている.よって,日本語構造は,英語で訳出が必要となる日本語ゼロ代名詞の位置情報と動詞によるそのゼロ代名詞となる格要素への意味的制約の情報が含んでいる.この意味的制約には,我々の独自の約3000種類からなる意味属性体系[6]の属性値を用いている.

例えば,日英対訳文(1)の日本語文からは次のような日本語構文意味構造が生成される.

```
(2) u-sent-1
    時制 : 過去, 完了相
    様相 : たい (希望)
    用言意味属性 : ガ格の身体動作, ガ格の思考動作
    ¦--- PRED : pred-1
    ¦          主動詞 : 読む (read)
    ¦--- CASE : case-1
    ¦     ¦    格関係 : 目的語
    ¦     ¦    助詞   : を
    ¦     ¦---- NP : np-1
    ¦     ¦       ¦-- N : 本
    ¦--- CASE : case-1
          ¦    格関係 : 主語
          ¦    助詞   : が
          ¦---- NP : φ-1 (意味制約 : 人間)
```

4. 2 英語文解析処理

英語文は,Brillの英語tagger[18]及び個々の単語の品詞情報を用いた部分構文解析系により解析する.英語文解析は以下の手順で行われる.

[Step 1] 英語文中の単語への品詞付け

Brillの英語 taggerを用いて,英語文中の各単語には品詞情報が付与される.

[Step 2] 英語文の部分構文解析

本処理では,まず,英語文中の名詞句及び述部が各単語の品詞情報を用いて認知される.ここでの述部は,動詞と様相・時制・相表現からなる連続する単語列のことを示し,個々の述部は態で区別される.次に,各述部の主語と直接目的語が英語文内の名詞句から認定される.主語は,述部に直接先行する名詞句,直接目的語は述部に直接後続する名詞句を認定する[2].

[Step-3] 英語文中の省略要素(ellipsis)の補完

英語文中の省略箇所の認定と補完が部分構文解析構造を用いて行われる.現時点では,等位接続語をはさんだ単文の主語の省略のみを対象としている.具体的には,省略箇所の認定と補完には次の規則が用いられる

[省略箇所の認定]

IF 英文中のある述部が能動態
 & その述部に直接先行する要素が名詞句ではなく等位接続語
THEN その述部の主語が省略箇所

---

[2] 本段階では,構文解析系として一般的な英文パーザを利用することも可能である.しかし,通常英文パーザは誤った構造を含む複数の解析多義を生成するため,最適な構造を選択する必要がでてくる.また,以降の処理段階で述べるとおり,本手法で必要な構文情報は,本段階で得られる部分構文構造に全て含まれている.よって,本論文ではBrillの英語 taggerと部分構文解析系のみを利用した.

[補完要素の認定]

IF　ある述部の主語が省略箇所と認定
　& その述部に先行する別の述部の主語が存在するか，主語は省略されており補完要素が既に認定されている

THEN 先行する述部の主語が補完要素

[Step-4] 英語文中の照応表現の照応解析

代名詞や定名詞句等の照応表現の照応解析が英語文の部分構文構造を用いて行われる．本照応解析処理によって，日本語文中のゼロ代名詞の認定された指示対象が代名詞や定名詞句等の照応表現であっても，それらの照応表現の指示対象が認定される．そして，図1のシステム全体によって，認定された英語文中の照応表現の指示対象とその日本語文中の対訳表現の情報から，その日本語ゼロ代名詞の文内・文間照応解析規則を抽出することが出来る[3]．

例えば，日英対訳文(1)の英語文からは次のような英語部分構文構造が生成される．

```
(3) u-sent-1
    ¦--- PRED : pred-1
    ¦          "want" : verb, non-3rd, sing. present.
    ¦          "to"   : to
    ¦          "read" : verb, base from
    ¦--- CASE : case-1
    ¦       ¦   格関係 : subject
    ¦       ¦---- NP : np-1
    ¦       ¦       ¦-- N : "I" : personal pronoun
    ¦--- CASE : case-1
    ¦       ¦   格関係 : direct object
    ¦       ¦---- NP : np-2
    ¦               ¦- "a"    : determiner
    ¦               "book" : noun, singular or mass
```

4．3　ゼロ代名詞と指示対象の認定

日本語文中のゼロ代名詞と英語文中の指示対象の認定処理は，以下の手順で行われる．本処理で用いる情報は，日本語文の構文意味構造と英語文の部分構文構造である．

[Step-1] 日英対訳表現対の認定

対訳関係にある日英表現対を[19]の提案手法で，次の情報を用いて認定する．

- 日英機械翻訳システム ALT-J/E の日英対訳辞書：本辞書は対訳関係にある表現対の認定に利用する．
- ALT-J/E の英文生成で用いる英語辞書：本辞書は語尾の違いを吸収するために利用する（例えば，対訳辞書エントリの英単語が"ing"型で英語文中の英単語が"ing"型でない場合）

なお，本処理では，前置詞や定冠詞などの機能語を取り除いて対訳関係にある表現対を認定する．これは，機能語は名詞句における主名詞や単文における動詞等のヘッドの種類に応じて変化する場合が多く対訳関係の認定が困難なためである．

本処理の結果，日英の構造中に対訳表現に関する情報が付与される．また，日本語解析の結果日本語表現に付与された意味情報が，それと対訳関係にある英語表現にも付与される．

[Step-2] 英語文中の日本語ゼロ代名詞の指示対象候補の認定

以下に示す英語表現が英語文中の指示対象候補として認定される．

- "I"や"you"等の人称代名詞
- "one"
- "it"や"that"等の指示詞
- 定冠詞を伴う名詞句等の定名詞句 (e.g. "the company")

[Step-3] 日本語文中のゼロ代名詞と英語文中の指示対象の対訳表現対の認定

対訳関係にある日英表現対と，日本語文中のゼロ代名詞－英語文中の指示対象の対訳表現対が，Step-2 で認定された日英対訳表現対候補，Step-2 で認定された英語文中の指示対象候補を用いて認定される．この決定は，候補がどれだけ強く日英で関係しているか，どれだけ多くの表現対が認定できるかに基づいて行われる．

日本語文中のゼロ代名詞と英語文中の指示対象の対訳表現対の認定規則は以下のとおりである．各規則は3章で用いた試験文を調査して人手で抽出したものである[4]．

以下に示す規則を各ゼロ代名詞に適用する際には，規則がマッチして指示対象が決定出来ても，指示対象候補がそのゼロ代名詞の意味的制約にマッチしない場合には，規則は適用されず，その指示対象候補は選ばれない．各規則は，以下の規則順で適用される．また，1種類の規則が1つのゼロ代名詞にマッチすると，規則9を

---

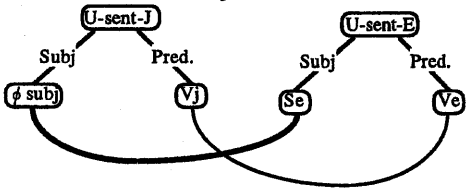
[3] 本段階は，現時点ではシステム中に実装していない．これは，高精度な英語文の照応解析を達成するためには，詳細な構文意味情報が必要となるからである．よって，本論文では，本段階を考慮した十分な検討を行なっていない．

[4] 規則の抽出過程では，比較的単純な日英の構文情報を用いて自動的に認定できるような規則を作るように配慮した．また，個々の規則がカバーする表現が広くなるようにも配慮した．

除いて，以降の別の規則は適用されない．

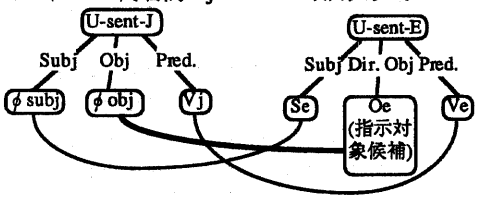
規則 1[5]

IF 日本語動詞(Vj)と英語動詞(Ve)が対訳関係
& Vjの主語相当[6](Sj)は対訳表現が未認定のゼロ代名詞
& Veの主語(Se)は対訳表現が未認定
THEN ゼロ代名詞 Sj は Se と対訳表現対



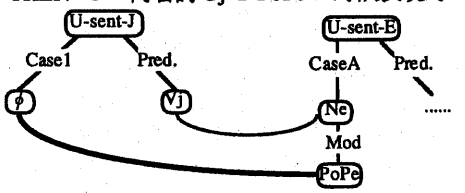
規則 2

IF 日本語動詞(Vj)と英語動詞(Ve)が対訳関係
& Vj の主語相当(Sj)はゼロ代名詞
& ゼロ代名詞 Sj と Ve の主語(Se)は対訳関係
& Vjの目的語相当[7](Oj)は対訳表現が未認定のゼロ代名詞
& Ve の直接目的語(Oe)は指示対象候補
& Oe は対訳表現が未認定
THEN ゼロ代名詞 Oj は Oe と対訳表現対



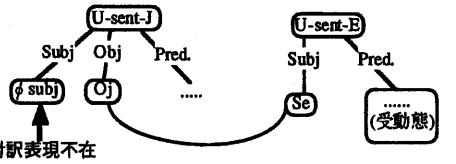
規則 3

IF 日本語動詞(Vj)と英語動詞の派生語である英語名詞(Ne)が対訳関係
& Ne は所有代名詞(PoPe)に修飾されている
& Vj のある格要素(Cj)は対訳表現が未認定のゼロ代名詞
& PoPe は対訳表現が未認定
THEN ゼロ代名詞 Cj と PoPe は対訳表現対



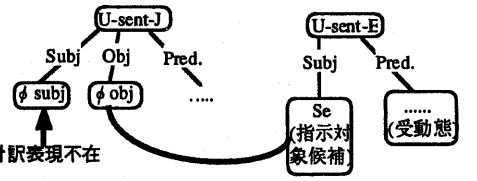
規則 4

IF 日本語動詞(Vj)の主語相当(Sj)は対訳表現が未認定のゼロ代名詞
& Vj の目的語相当(Oj)は英文中の主語格要素(Se)と対訳関係
& Se の述部が受動態
THEN Sj は対訳表現不在と認定（受動態で翻訳されているため，ゼロ代名詞 Sj には英語文中に指示対象が存在しない）



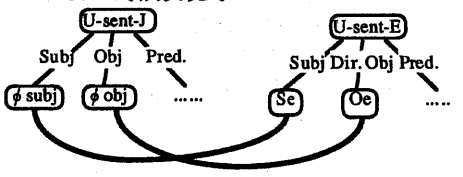
規則 5

IF 日本語動詞(Vj)の主語相当(Sj)は対訳表現が未認定のゼロ代名詞
& Vj の目的語相当(Oj)は対訳表現が未認定のゼロ代名詞
& 英文中のある述部(Pe)が受動態
& Pe の主語(Se)が指示対象候補
THEN ゼロ代名詞 Sj は対訳表現不在と認定
& ゼロ代名詞 Oj と Se は対訳表現対
（英語文が受動態で翻訳されているため）



規則 6

IF 日本語文中のある日本語単位文(U-sent-J)の日本語動詞(Vj)の主語相当(Sj)と目的語相当(Oj)は対訳表現が未認定のゼロ代名詞
& 英語文中のある英語単位文(U-sent-E)の主語(Se)と目的語(So)は対訳表現が日本語文中の要素又は日本語文中の U-sent-J 中の要素と未認定
THEN ゼロ代名詞 Sj と Se 及びゼロ代名詞 Oj と Oe は対訳表現対



規則 7

(1) IF 対訳表現が未認定のゼロ代名詞(Cj)が日本語文中に１つしかない

---

5 規則説明図中で細線は認定済みの対訳表現対を，太線は本規則で認定された対訳表現対を示す．

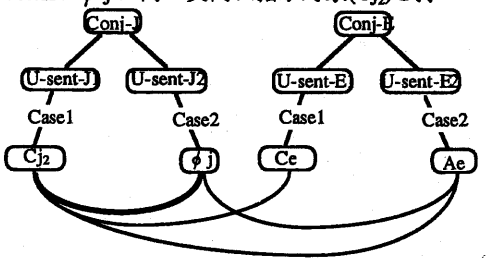
6 主語相当とは，ガ格要素か，MT システムで英語主語に翻訳される日本語格要素のことを示す．

7 目的語相当とは，を格要素か，MT システムで英語直接目的語に翻訳される日本語格要素のことを示す．

& 対訳表現が未認定の指示対象候補(Aei)が英語文中に1つ以上存在する
THEN ゼロ代名詞 Cj は次の優先度に従い Aei より対訳表現を1つ決定
人称代名詞 > "one" > 決定詞 > 定名詞句

規則8
IF 対訳表現が未認定のゼロ代名詞(Cj)がある
& 全ての指示対象候補は対訳表現が認定済
& Cj の単位文内に対訳表現が存在しない指示対象候補(Aei)が1つ以上存在する
THEN ゼロ代名詞 Cj は次の優先度に従い Aei より対訳表現を1つ決定
人称代名詞 > "one" > 決定詞 > 定名詞句

規則9
IF 規則1-8で決定されたゼロ代名詞(ϕ j)の対訳表現(Ae)は日本語文中に別の対訳表現(Cj_2)も認定されている
THEN ϕ j は同一文内に指示対象(Cj_2)を持つ



規則10
IF 対訳表現の未認定のゼロ代名詞が残る
THEN これらのゼロ代名詞は先行詞が英文中に明示されておらず対訳表現不在と認定

以上の規則によると,例文(1)のゼロ代名詞には規則1が適用され,"I"が指示対象として認定される.

5. 評価

5. 1 評価方法

4章で述べた,日英対訳コーパスからゼロ代名詞とその指示対象を自動抽出する手法を,日英対訳文に適用してその性能を評価する.個々の評価条件は以下のとおりである.

5. 1. 1 評価の対象対訳文

3章の検討で用いた日英機械翻訳システム評価用例文3718文の文対応済みの日英対訳例文集のなかでゼロ代名詞が存在する文対(554件;437文対)を評価の対象とした.

5. 1. 2 日本語文・英語文の解析

ゼロ代名詞が存在する日英対訳文に対して,日英機械翻訳システム ALT-J/E を日本語解析系に利用して日本語文の構文意味構造を生成し,Brillの英語taggerと4. 2節で述べた英文部分構文解析系を英語解析系に利用して英語文の部分構文構造を生成した.これら解析系の解析失敗による手法全体の評価への影響を避けるため,日英の解析失敗は人手で修正して評価した.また,英語文中の照応表現の照応解析による性能への影響を検証するために,英語照応表現の指示対象を人手で認定した場合の性能も評価した.

5. 1. 3 指示対象

ゼロ代名詞を含む日英対訳文対に対して各ゼロ代名詞の指示対象を4章の手法で自動認定した.

5. 1. 4 評価項目

以下の3種類のゼロ代名詞別に指示対象の認定精度を評価した.

[type A] 文内照応ゼロ代名詞 (150件)
[type B] 英語文で受動態で訳されるため訳出不要となった文章外照応ゼロ代名詞 (157件)
[type C] 英語文中で明示的に訳されている文章外照応ゼロ代名詞 (247件)

5. 1. 5 認定成功条件

4. 3節で述べた規則によって,type A, Cのゼロ代名詞の指示対象が正しく認定できるか,type Bのゼロ代名詞が対訳表現不在と認定される場合を正解とした.

5. 2 評価結果

ゼロ代名詞の指示対象の認定精度を表2に示す.この表から,3タイプのゼロ代名詞に対して4. 3節の規則を用いて指示対象を認定した場合,英語の照応解析を行った場合で91.5%,行わなかった場合でも87.2%という認定精度を得た.特に,英語文で訳出不要となった文章外照応ゼロ代名詞(type B)の認定精度は両評価条件とも100%という高い精度を達成した.さらに,英語の照応解析は文内照応ゼロ代名詞(type A)の指示対象認定精度にのみ影響をあたえることが分かった(89.3%(照応解析あり)と73.3%(照応解析なし)).以上の結果から,たとえ英語照応解析処理を行わない場合でも,本手法は比較的高い精度で文内照応ゼロ代名詞の指示対象が正しく認定できることが分かった.

本論文の提案手法で指示対象が正しく認定出来なかったゼロ代名詞(47件)を詳細に分析した結果を表3に示す.最も多い認定失敗原因は,英語が意訳されており,指示対象に相当する表現が英語文中に存在しない場合である(47件中42件).これらの対訳文対は,人間訳では指示対象に相当

表2　ゼロ代名詞の指示対象認定精度

英文解析条件	文内照応	文章外照応		小計
		受動態	明示訳出	
照応解析あり	89.3% (134/150)	100.0% (157/157)	87.4% (216/247)	91.5% (507/554)
照応解析なし	73.3% (110/150)	100.0% (157/157)	87.4% (216/247)	87.2% (483/554)

表3　指示対象の認定失敗原因

認定失敗原因	文内照応	文章外照応 (明示訳出)	計
意訳	11	31	42
受動態	4	0	4
対訳認定失敗	1	0	1

する表現が不要でも，機械翻訳システムは人間の様に意訳が出来ないためゼロ代名詞の照応解析が必要となる場合である．よって，この認定失敗は，提案手法の方式限界によるものといえる．また残りの5件は，全て文内照応ゼロ代名詞で起こった．この5件中4件は，受動態で英訳されているため指示対象は英語文中に存在しないが，人間は指示対象が同一文内にあると容易に認定できる場合である．この解析失敗もやはり提案手法の方式限界であるによるものである．残りの1件は，ゼロ代名詞の動詞と指示対象の動詞の対訳関係が認定できず，指示対象が認定された場合である．これは方式限界ではなく，対訳表現対抽出アルゴリズムの問題である．

以上のことから，本論文で用いた試験文に関しては，554件のゼロ代名詞中508件は提案手法により方式的に指示対象を認定することが出来，そのほとんどが提案手法によって正しく指示対象を認定することが出来た(99.8%; 508件中507件)．以上の結果から，本論文で提案した手法の有効性を示すことが出来た．

4．まとめ

本稿では，日英対訳コーパスを用いた，日本語文内のゼロ代名詞と英語文内の指示対象を自動的に認定する手法を提案した．本論文では対訳文対として日英のみを用いた．しかし，提案手法は，英語イタリア語対等様々な言語対でも実現可能であり，提案手法の認定精度は両原語の表現の異なり具合に依存することが予想される．

今後は，文間照応ゼロ代名詞に対する提案手法の有効性を評価するとともに，より大規模な対訳コーパスに対しても性能評価を行いたい．また，英語文解析に汎用的な英語構文解析系を用いた手法も検討していきたい．さらに，未知語等ノイズの多いテキストに対してもより有効に働くより強力な対訳表現対抽出アルゴリズムや，対訳コーパスからの対訳文対対抽出アルゴリズムを提案手法と結合させた検討も行っていきたい．

謝辞

1995年から1996年までのマンチェスタ理工科大学(UMIST)滞在中，本技術に関して貴重な議論をしていただいた辻井潤一教授に感謝致します．

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 (1994).

[5] 中岩,池原：日英の構文的対応関係に着目した日本語用言意味属性の分類，情報処理学会論文誌，Vol.38 No.2 (1997).

[6] 池原ら編: 日本語語彙大系，岩波書店 (1997).

[7] 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 (1993).

[8] 中岩,池原：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，自然言語処理，Vol 3 No.4 (1996).

[9] 中岩，白井，池原：日英機械翻訳における語用論的・意味論的制約を用いたゼロ代名詞の文章外照応解析，情報処理学会論文誌，Vol.38, No.11 (1997).

[10] Nasukawa, T.: Full-text processing: improving a practical NLP system based on surface information within the context, Proc of COLING-96 (1996).

[11] 村田，長尾: 用例や表層表現を用いた日本語文章中の指示詞・代名詞・ゼロ代名詞の指示対象の推定，自然言語処理，Vol.4, No.1 (1997).

[12] Melamed, I. D.: Automatic Detection of Omissions in Translations, Proc of COLING-96 (1996).

[13] Nakaiwa, H.: Automatic Extraction of Rules for Anaphora Resolution of Japanese Zero Pronouns from Aligned Sentence Paris, Proc. of ACL/EACL-97 workshop on Operational Factors in Practical, Robust, Anaphora Resolution for Unrestricted Texts (1997).

[14] 中岩：日英対訳コーパスを用いた文章外照応ゼロ代名詞解析規則の自動獲得，情報処理学会第55回全国大会講演論文集，2-45 (1997).

[15] 池原,白井,小倉：言語表現体系の違いに着目した日英機械翻訳機能試験項目の構成，人工知能学会誌論文，Vol.9, No.4 (1994).

[16] 池原他：言語における話者の認識と多段変換方式，情報処理学会論文誌，Vol.28, No. 12 (1987).

[17] 八巻他：特集論文 日英機械翻訳技術，NTT R&D Vol. 46 No. 12 (1997).

[18] Brill, E. : A simple rule-based part of speech tagger, Proc. of ANLP'92 (1992).

[19] Yamada, S. et. al. : A new method of automatically aligning expressions within aligned sentence pairs, Proc. of NeMLaP2 (1996).