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

4J－1

中岩浩巳
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．特に，日本語では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，日英機械翻訳システムにおいては，この省略格要素（ゼロ代名詞）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いるが，翻訳対象分野を限定しない機械翻訳への応用を考えると，解析精度，対象とする言語現象，必要となる知識量の面で問題があった．これに対しては，分類した用言の意味属性[1]，様相表現，接続表現を用いて指示対象を決定する機械翻訳に適した照応解析手法が提案されている[2][3][4]．

しかしこれら従来手法では，基本的に人間が照応解析規則を作成する必要があるため，網羅的な照応解析規則の作成には，かなりの専門知識と労力が必要となる．さらに，解析対象分野に応じて，指示対象の傾向の異なるゼロ代名詞が存在するので，分野に依存した規則を作成する必要がでてくるが，分野毎に規則を作成することはその労力を考えると不可能である．よって，このゼロ代名詞の照応解析ルールを効率的に獲得する手法の実現が望まれている．

近年，既存のコーパスを用いて，コーパス中の言語現象を分析し，その結果を基に照応解析規則を抽出する手法が提案されている[5],[6]．しかし，それらの手法は解析対象言語のコーパスのみを使用するため，その言語ではほぼ常にゼロ化される要素への規則を抽出することは困難である．よって，利用するコーパスとしては，解析対象の言語と他の言語の対訳コーパス利用することが有望である，特に，日本語と英語のように言語族が異なる場合には，省略現象の傾向が異なるため，ある言語の文ではゼロ化されている要素が，その文の対訳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．

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，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，文章外照応ゼロ代名詞の照応解析規則を抽出する手法を提案する．

２．システム構成

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指示対象を抽出し，その結果をもとに，日本語ゼロ代名詞の照応解析規則を自動生成するシステムの構成図を図１に示す．図の通り，入力された日英対訳コーパス中の対訳関係にある日本語文と英語文を解析し，その文対から対訳関係にある表現対を抽出する[7]．次に，日本語ゼロ代名詞および英語文中の指示対象を抽出する[8],[9]．そして，英語指示対象をもとに日本語指示対象を抽出する．以上の結果から，日本語解析結果を参考に，ゼロ代名詞照応解析規則を作成する．その後は，同じ対訳コーパスを対象にその新規作成された規則を用いて再度照応解析し，新規作成された規則の有効性を検証しつつ，この学習過程を繰り返す．本システムは，日英機械翻訳システム ALT-J/E 中に実装中である．

３．ゼロ代名詞照応解析規則の生成処理

ここでは，図１のシステム全体の内，日本語構文意味構造とその構造中から認定されたゼロ代名詞及びその指示対象を用いて，照応解析規則を生成する処理について述べる．規則の生成には，日本語構文意味構造中の用言意味属性，様相表現，接続語を照

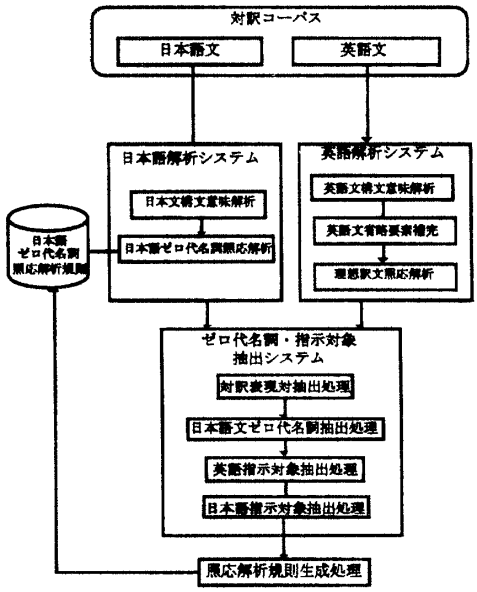

図１　照応解析規則自動抽出処理の構成図

Automatic Extraction of Rules for Anaphora Resolution of Japanese Zero Pronoun with Deictic Reference from Aligned Sentence Paris.
Hiromi Nakaiwa, NTT Communication Science Labs.

応解析条件として用いる．ALT-J/E に実装する際には，個々のゼロ代名詞に対して，ゼロ代名詞を含む単位文中の用言意味属性(107 カテゴリ)及び様相属性(134 カテゴリ)と，ゼロ代名詞を含む単位文に隣接する接続属性(56カテゴリ)を用いて規則を作成する．

同じ属性からなる解析条件に複数の指示対象が現れた場合は，同じ条件中最も頻度の高い指示対象をその解析条件で決める指示対象としてルール化する．例えば，可能の様相表現を伴うガ格のゼロ代名詞が8件存在し，その内5件が'I'を指示し，3 件が'you'を指示する場合，次の規則が抽出される．

規則例　IF　　が格＝φ　and 様相表現=可能
　　THEN　指示対象＝'I'

4．評価

4．1　評価方法

本論文で提案した照応解析規則の自動獲得手法は，日英対訳コーパスから照応解析規則を自動的に抽出し，その規則の解析精度を調査することで評価した．評価条件の詳細は以下のとおりである．

獲得対象：日英機械翻訳システム評価用対訳文集[10]中の文章外照応ゼロ代名詞(371件)を含む文．

照応解析規則：上記371 件のゼロ代名詞に対し，日本語構文意味構造中の用言意味属性，様相表現，接続語の３種類の属性値を用いて３節で述べた方法で規則を作成する．なお，本評価では，用言による格への意味的制約は使用しなかった．

評価項目：解析規則の種類と解析精度の関係を調べるため，様相表現，用言意味属性，接続語を用いて自動的に作成した規則の解析精度を検証した．また，解析精度の絶対評価を行うため，１種類の格に出現するゼロ代名詞において，最も出現頻度の高いものを指示対象とした場合の解析精度も調査した．なお，個々の評価では，以下の２種類の方法で解析精度を評価した．

- ウインドウテスト　371 件すべてのゼロ代名詞を規則獲得のために利用し，解析精度を評価．
- ブラインドテスト　370 件のゼロ代名詞を用いて規則獲得し，残りの 1 件に適用する過程を371 回繰り返し評価．

解析成功条件：指示対象を正しく決定する規則が，抽出規則中に含まれる場合を成功とする．これは，有効な規則を生成するための必要条件である．

4．2　評価結果

獲得した照応解析規則の解析精度を表１に示す．この表から，３種類全ての制約を用いて自動作成した場合，ウインドウテストで 99.2%，ブラインドテストで 87.6%と，頻度のみを用いた場合の約 46%にくらべかなり高い精度が得られることが分かった．

表１　照応解析条件と解析精度の関係

条件	解析精度	
	ウインドウテスト	ブラインドテスト
様相表現のみ	74.9% (278 件)	64.2% (238 件)
用言意味属性のみ	70.9% (263 件)	52.6% (195 件)
接続語のみ	55.0% (204 件)	48.8% (181 件)
様相表現と用言意味属性	95.2% (353 件)	81.7% (303 件)
様相表現と接続語	90.3% (335 件)	79.5% (295 件)
用言意味属性と接続語	87.7% (326 件)	68.2% (253 件)
様相表現,用言意味属性と接続語	99.2% (368 件)	87.6% (325 件)
頻度情報のみ	46.4% (172 件)	46.1% (171 件)

また，個々の条件を単独で用いた場合は，様相表現，用言意味属性，接続語の順で，文章外に指示対象を持つゼロ代名詞の照応解析に有効になることが分かった．以上の結果から提案した照応解析規則の自動獲得手法の有効性を示すことが出来た．

5．まとめ

本稿では，日英対訳コーパスを用いた，文章外照応ゼロ代名詞の照応解析規則を自動的に抽出する手法を提案した．今後は，文内・文間照応ゼロ代名詞に関しても自動獲得した規則の評価を行うとともに，効果的な規則の抽出法の検討を行いたい．

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] 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).

[2] 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).

[3] 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).

[4] Nakaiwa, H. et al : Resolution of Japanese Zero Pronouns with Deictic Reference, Proc of COLING'96 (1996).

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

[6] 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).

[7] Yamada, S. et al : A New Method of Automatically Aligning Expressions within Aligned Sentence Pairs, Proc. of NeMLaP2 (1996).

[8] 中岩，山田: 日英対訳コーパスからのゼロ代名詞とその指示対象の自動抽出，言語処理学会第３回年次大会，(1997).

[9] Nakaiwa, H.: Automatic Identification of Zero Pronouns and their Antecedents within Aligned Sentence Pairs, Proc of 5th workshop on Very Large Corpora (1997).

[10] 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,Vol.9, No.5 (1994).