日本語生成に於ける対話文脈構造と代名詞省略

6B−7

赤峯享* 亀井真一郎* 飯田仁**
(*NEC C&C 情報研究所 **ATR 自動通訳電話研究所)

1. はじめに

日本語文では、必須格等の文の骨格をなす要素でも既知の情報であれば、省略して表現される。特に対話においては、効率的な情報交換を行う必要があるため、話し手同士が共通に了解している事柄を明示しない。このことが、機械翻訳等の自然言語処理において、大きな問題となっている。

対話文を英語から日本語へ翻訳する場合、英文に存在する単語をそのまま全て日本語として表出してしまうと、冗長で可読性の悪い文になり、必要な情報が伝わり難くなってしまう。特に英語で必須格を埋めるために用いられる代名詞は、日本語として表出すると不自然な可読性の悪い文になる場合が多い。これは、英語と日本語の文法の差を考慮せずに翻訳した結果、聞き手に自明である要素を表出してしまい、聞き手がその要素に特別な意味づけをしてしまうためだと思われる。

ある要素が聞き手にとって自明となるのは、1) 文脈(前からの文の流れ)による場合と、2) 一文内での他の要素との対応による場合がある。日本語生成における代名詞の省略は上記の2つ条件を考慮して行うべきであるが、今回は、まず代名詞の省略の現象を明らかにするために、現状の機械翻訳システムに新しい生成部を組み込んで実際に評価することができる2) について考察を行った。

本稿では、一文内に表れるモダリティと文の主体との対応関係に着目することで、代名詞が省略可能であることを示す。さらに、これを取り入れて、代名詞を省略し、自然な日本文を生成する方式について述べる。また、実際にこの方式を英日機械翻訳システムに組み込み、対話文コーパスの翻訳に適用した結果を示す。

2. 省略の基本方針

例1では、英語の表層に現れる単語を全て日本語に翻訳すると、対話文としては冗長な表現になる。

<例1>
原文: I 'd like to ask your phone number.
翻訳文: 私はあなたの電話番号を尋ねたいのですが。

英語の原文は冗長性のない文であったはずである。しかしながら、日本語として冗長であるということは、翻訳の過程で日本語として不必要なものを生成したことを意味する。

生成の立場から省略を考えた場合、適切な省略を行うためには、何が、どういう場合に省略可能であるかを明確にする必要がある。省略の根本原則は、「省略されるべき要素は、言語的、或は非言語的文脈から、復元可能でなければならない[1]」ということである。つまり、生成において省略を行うためには、その要素が復元可能であることが必要条件である。また、対話文おいては、冗長な表現はできるだけ避けるはずである。したがって、上記の必要条件を満たせば、例外的に省略できない場合以外は、省略を行うことが可能であると考える。そこで、以下の方針で代名詞の省略を行う。

「ある要素が他の要素を手がかりとして復元可能である場合に省略を行う。」
ただし、以下のものは例外として省略を行わない。
- 焦点が当たっているもの。
- 日本語で文法的に必要なもの。

また、評価を的確に行うことを可能にするために、今回は以下のものに絞って考察を行う。

- 対話文の条件:登場人物は話し手と聞き手のみ(「私(私達)」、「あなた(あなた達)」)。
- 省略の条件:文脈情報を考慮せずに一文単位で復元可能性が保証できるもの。

3. 文構造と省略要素

3.1. 文末のモダリティと主体との対応

日本語文の表層の基本構造は、図1のようになっており、モダリティと文の主体の間には依存関係があることが知られている[2]。この表層の構造は、深層構造より生成されたと考えられる[3]。

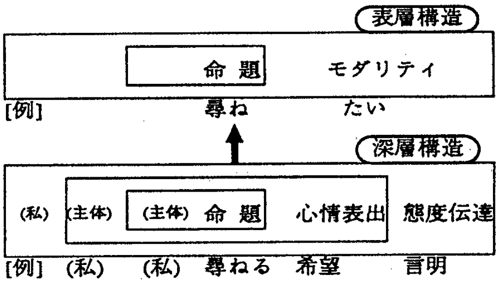

図1: 文の構造

例えば、例1の翻訳文ではモダリティは「たい」であり、「たい」は深層構造では、図1の例のようになる。表層のモダリティの形が「たい」の言い切りとなるのは、命題の主体と心情表出の主体が共に「私」の場合だけである。つまり、「たい」を生成することによって、文の主体は聞き手にとって復元可能であるので、「私は」を省略して、文を生成することができる。同様の理由により、依頼や意志等の他のモダリティについても主体を省略することが可能である。モダリティと復元可能な主体の代名詞の対応例を図2に示す。

3.2. 授与動詞と授与の相手との対応

動詞が持つ特徴を利用して復元可能なものとして、「与える」、「教える」等の相手側へのものの移動を表す授与

Pronoun ellipsis in Japanese generation
Susumu Akamine*,Shin-ichiro Kamei*, Hitoshi Iida**
*NEC Corp. , **ATR Interpreting Telephony Research Labs.

心情表出	態度伝達	命題の主体	表出の主体	表層のモダリティ
希望	言明	私	私	たい(希望)
		あなた	私	もらいたい(依頼) 下さい(命令) よい(許可)
	疑問 (確認)	あなた	あなた	たい + か (たい + ね)
		私	あなた	もらいたい+か よい+ か (よい + ね)
意志	言明	私	私	つもりだ(意志)
	疑問 (確認)	あなた	あなた	つもり + か (つもり + ね)

図 2: モダリティと復元可能な主体の対応例

動詞がある。これらの動詞は、「(主体) が+ (相手) に+ (対象) を+動詞」の構文で用いられる。2者間の対話の場合は、主体が決まれば、相手は省略しても復元可能である。つまり、主体が「あなた」の場合、相手の「私」が、主体が「私」の場合、相手の「あなた」が省略可能である。

3.3. 名詞の尊敬表現と所有との対応

「あなたの電話番号」のように、所有格の「あなた」についても、話し手と聞き手の関係により復元可能性を保証することができる場合がある。名詞 (「電話番号」) を尊敬表現にして、「お電話番号」と生成することで、「あなた」を省略することができる。

4. 生成方式

英日翻訳システム [4, 5, 6] の日本語生成部にこの方式を組み込んでＡＴＲコーパス (「国際会議の参加申込みに関する問い合わせ」、約３００文) に対して評価を行った。このシステムは、中間言語方式で翻訳を行っており、生成の入力となる概念表現は、モダリティをキーとして適切な省略を行うために以下のような特徴を持つ。

- 言語固有の文構造には依存しない。
- 命題とモダリティが分離して表される。
  (命題の部分のみ依存構造として表される)
- 命題部分は概念とその関係によって示される。

省略を行うことで通常の生成に副作用を与えないことを保証し、他の対応関係についても容易に拡張を行うことを可能とするためために、概念表現より、以下の方式で生成を行った。例 1 の文の生成過程を図 3 に示す。

- 概念表現で、省略の可能性をチェック：
  モダリティとの対応等により、省略候補のマーク。
- 構文木上で、省略要素を確定：
  対応する助動詞、接頭の生成を確定した際に、省略候補の省略可能性を確定。
- 省略可能な要素をスキップして出力。

5. 生成結果

このシステムを用いて、実際に約 300 文のコーパスに対して翻訳を行った結果、主体を表す「私」、「あなた」については、過剰な省略を行うことなしに自然な日本文の生成

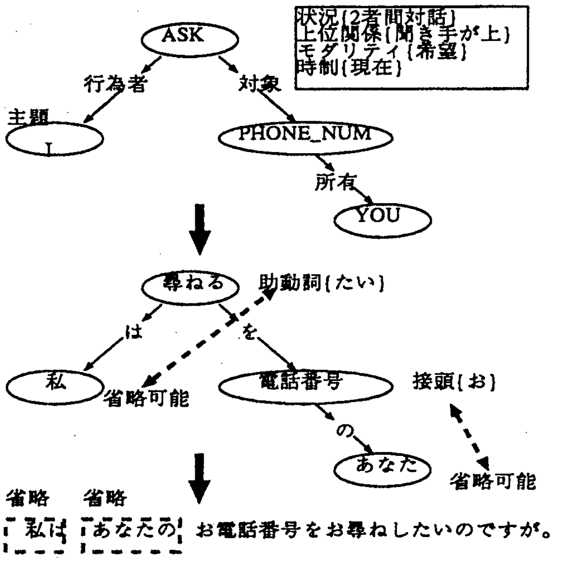

図 3: 生成方式

が可能であった。ただし、今回の省略の対象外であった、it 等の文脈を把握して省略 (もしくは、反復) を行う必要のあるもの、some,any 等の数量詞については冗長な不自然な表現が生成されたものがあった。

原文： May I ask your name? …許可、名詞の尊敬
翻訳文：(私は)(あなたの) お名前をお尋ねしてもよろしいですか。

原文： Can you tell me some information about the conference? … 依頼、授与動詞
翻訳文：(あなたは)(私に) 会議についてのいくらかの情報を教えて下さいますか。

6. おわりに

英日の対話文の翻訳において、モダリティ等をキーにして代名詞の省略を行うことで自然な日本語文の生成を行う方式を示した。また、この方式を実際の英日翻訳システムに実装し、約 300 文の ATR の対話コーパスに対して評価を行った。その結果、本方式の有効性を確認することができた。

今回は、一文単位の限定された代名詞の省略だけをおこなったが、今後は、文脈を考慮した省略についても評価を行う予定である。また、これらにより得られた省略可能性を、解析における省略要素の補完の手がかりとして用いる予定である。

参考文献

[1] 久野すすむ,「談話の文法」, 大修館書店, 1978

[2] 仁田義雄,「日本語のモダリティ」, くろしお出版,1989

[3] 林四郎,「談話の研究と教育 I」, 国立国語研究所,pp43-62,1983

[4] 土井他,「用例に基づいて英語前置詞句の訳し分けを行う英日翻訳システム」, 第 46 回情処全国大会,5B-1,1993

[5] 隅田他,「用例に基づいて英語前置詞句の係り先決定を行う英日翻訳システム」, 第 46 回情処全国大会,5B-2,1993

[6] 石原他,「対話文の英日機会翻訳における日本語待遇表現の生成」, 第 46 回情処全国大会,1993