

語順と省略に着目した自然な日本語対話文の生成*

4K-1

赤峯享 古瀬蔵†

ATR 音声翻訳通信研究所‡

e-mail: {akamine, furuse}@itl.atr.co.jp

1 はじめに

従来、機械翻訳や対話システム等で文を生成する場合、「文法的」に正しい文を生成することが最優先の目的とされてきた。その結果、生成された文を人間が読んだ場合、必ずしも自然な、理解しやすい文になっているとは限らなかった。同一の意味内容をもつ文の中から、最も自然な日本語文を生成しようとする試みが、最近なされてはきている [1, 2, 3] が、これらは書き言葉を対象としたもので、話し言葉を対象にしたものは、ほとんど報告されていない。

本稿では、瞬間的に生じては消える一回的なものである話し言葉（音声言語）を対象とし、自然な日本語対話文を生成するための語順と省略に関して述べる。まず、自然な対話文を生成するための基本方針を示す。次に語順や省略についての日本語文の特色について考察する。最後に日本語文の特色にしたがって自然な文を生成する方式について述べる。

2 自然な対話文生成の方針

文生成（翻訳）対象を、「ホテルを予約する場合の対話」のように、話し手と聞き手の間で、対話の目的が明確であるものとする。この場合、対話を行なう目的は、情報交換を行なうことであり、効率的な情報交換ができる文を自然な文とする。つまり、自然な日本語文とは、「一度聞いただけで、話者の意図や言いたいことを確実に伝達できる文」である。そこで、基本的に以下の方針を元にして、自然な対話文を生成することを考える。

方針1 まず重要な表現を先に生成し、重要度の低い表現を付加的に付け加える形で生成する。

方針2 冗長な表現を省略し、できる限り短い文を生成する。

方針3 係り受け等の曖昧さのない文を生成する。

例えば、従来の文生成では、倒置された特殊な文として決して生成されることがなかった下の (I) ような文においても、

(I) 「3000円かかります。京都駅から会議の会場へタクシーで行くと」

聞き手にとって重要な情報（未知情報）が「3000円かかる」ということのみであれば、1) 聞き手にとって重要な情報をまず生成する、2) 倒置された文において係り受けなどの曖昧さが生じることがない、とう条件で (I) の文が自然であるときとみなし、倒置された文をも生成する。

*A Japanese generation based on word ordering and ellipsis

†Susumu AKAMINE Osamu FURUSE

‡ATR Interpreting Telecommunications Research Labs.

3 日本語文の特色

3.1 語順

日本語文は、比較的語順が自由であると言われるが、実際には、様々な制約が報告されている [2, 3]。まず、一文内の日本語文の語順の制約について整理する。

A. 述語内部

述語内部の助動詞の並びについては、固定しており、自由度はない。これは、書き言葉だけでなく、話し言葉においても同様である。その語順は以下のようになる。

動詞 → ボイス（「れる」）→ 授受（「てもらう」）
→ 心情表出（「たい」）→ 話者推量（「だろう」）
→ 態度伝達（「か」）

例) 「送ってもらいたいのだろうか。」

B. 修飾関係

書き言葉では、修飾語→被修飾語の順になり、自由度はない。しかしながら、話し言葉では、「(彼ではなくて)私が行きます。京都まで」のように倒置された文であっても不自然でないものが見られる。

修飾語→被修飾語

例) 「私が 京都まで 行きます」、「赤い 本」

C. 格成分

格要素については、一般的には以下の語順で生成されることが多く、この語順で生成することが自然だと言われるが、書き言葉においても語順の自由度が大きい。

時 → 場所 → 主格「～が」 → 相手「～に」
→ 対象「～を」 → 動詞

例) 「明日 東京で 私が 彼に 本を 渡す」

D. 主題

「は」等の提題を表す助詞がついて、主題となった格要素を上格成分の語順に反しても格成分の先頭に置く。

例) 「本は 明日 東京で 私が 彼に 渡す」

E. 長い成分

埋め込み句等の長い成分は、短い成分の前に置く。

例) 彼が昨日せっかく予約してくれたホテルを 彼女が キャンセルした

以上、述べてきた語順の制約は、「A. 述語内部」を除けば、どれも文法的に絶対的なものではない。つまり、「B. 修飾関係」、「C. 格成分」、「D. 主題」、「E. 長い成分」については、一般の場合に読み手に必要な情報を曖昧性を排除して効率的に伝えるための書き言葉におけるヒューリスティックである。B, E については、係り受けの曖昧性を避けるためのものであり、C, D につ

表 1: モダリティと冗長な主体との対応例

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)

表 2: 敬語表現と冗長な主体者と受け手

主体	受け手	種類	例
自分	相手	受手尊敬	お送りする
相手	自分	主体尊敬	送ってくださる
相手	—	主体尊敬	お送りになる

いては、本来、旧情報・新情報というような文脈情報がなければ解決できないものであると思われる。従って、旧情報・新情報や焦点といった文脈情報から聞き手にとって重要な表現が得られれば、方針 1, 3 に従って生成を行なうことで、話し言葉において適切な語順で生成することが可能である。

3.2 省略

日本語文では、必須格等の文の骨格をなす要素でも言語情報や文脈情報から復元可能であれば、省略して表現することができる [5]。

一文内の言語情報から省略可能になるものとしては、モダリティとの関係で主体が冗長となるものがある [6]。例えば、「私は電話番号を尋ねたいのですが。」という文において、希望を表すモダリティ「たいのですが」は、必ず、主体者として「私」を要求する。したがって、「私は」は、冗長な表現であり、「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者を省略することが可能である。表 1 にモダリティと冗長な主体者との対応例を示す。

同様に、対話での登場人物が話し手と聞き手の 2 者のみと特定できる場合、動詞の敬語表現により、主体者や受け手を省略することが可能である。例えば、「私があなたに本をお送りします」というような文では、「お送りする」という表現により「私が」、「あなたに」は、冗長な表現になり、省略して、文を生成することができる。表 2 に敬語表現と冗長な主体者と受け手との関係を示す。

4 生成方式

上記の日本語の特色を考慮して、語順や省略を取り扱い、自然な日本語文を生成する方式について以下に示す。また、図 1 に基本的な流れを示す。従来の生成では、「正しい文」を文法的な規則を適用することによって生成していたが、この方式では、まず、聞き手にとって必要な情報が認識できるような語順を決定し、次に生成結果が入力と同一であるかのチェックを行なうことで、「正しい文」を生成する。

生成方式:

入力は、日本語の語彙に係り受け等の情報が付加され、文脈情報から聞き手にとって未知である情報が与えられるとする。

A-1) まず、述語内部の語順、モダリティをキーにした省略等の固定された成分について語順の決定および省略を行なう。A-2) 未知情報、未知情報の係り先の順で語順を決定する。A-3) 残りの語順が未定の部分について、3章のヒューリスティックを用いて、語順を決定し、表現 1 として出力する。B) 表現 1 を解析し、係り受け構造や省略された要素の復元を行なう。C) 解析結果(表現 2) と入力との係り受け等が同一かどうかのチェックを行なう。同一の場合は、表現 1 を形態素生成を行ない、生成文を出力する。同一でない場合は、入力に対して、再び A-1), A-3)、形態素生成を行ない生成文を出力する。

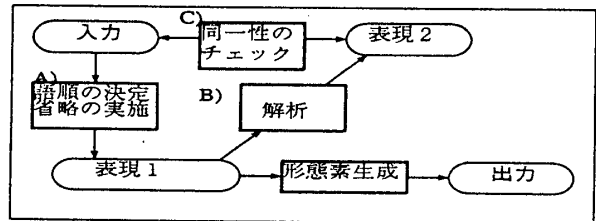


図 1: 生成方式

5 おわりに

自然な日本語対話文を生成するための語順と省略に関して考察を行ない、その生成方式を提案した。今後は、英日対話文翻訳システムにこの生成方式を組み込み、実際のシステムとして評価を行なう予定である。

参考文献

- [1] 中村他:「視点の理論に基づく文生成の試み」, 情報処理自然言語処理研究会 83-8, 1991
- [2] 乾他:「文章生成における推敲の役割」, 情報処理自然言語処理研究会 83-7, 1991
- [3] 野崎他:「自然な日本語文を生成するための語順」, 情報処理第 41 回全国大会, 1990
- [4] 寺村他編集:「ケーススタディ 日本文法」, 桜楓社, 1987
- [5] 久野すすむ, 「談話の文法」, 大修館書店, 1978
- [6] 赤峯他:「日本語生成に於ける対話文脈構造と代名詞省略」, 第 46 回情報処理全国大会, 1993