

照応・省略を含む日本語論説文生成

滑川裕樹[†] 乾健太郎^{††} 徳永健伸[†] 田中穂積[†][†]東京工業大学大学院情報理工学研究所 ^{††}九州工業大学知能情報工学科

1 はじめに

計算機によって日本語文章を生成させるとき、文脈上の個体を参照する照応・省略表現を適切に生成することが必要である。照応・省略表現の指示対象は読み手や聞き手が容易に推定できなければならないから、これらの表現を生成するためには現在の文脈を考慮しながら、どの要素に照応・省略表現を用いるか決定しなければならない。

本研究では、省略表現を適切に生成するための基礎的研究として、論説文を生成対象とし、省略要素を復元した発話ゴールから適切な日本語文章を生成することを試みた。

2 生成システムの説明

以下で本研究で用いた日本語文生成システムを説明する。

2.1 生成モデル

本研究の自然言語生成システムは、坂庭 [5] で提案されている 3 層モデルを用いている。

自然言語生成モデルとしては、内容決定部と言語表現決定部からモデルを構成する 2 層モデルがある。3 層モデルでは、2 層モデルの言語表現決定部を発話プラン生成部と表層生成部に分割し、発話プラン生成部に領域依存性を吸収させることにより領域独立な表層生成部を実現している。システム全体の構成を図 2 に示す。

2.2 システムの動作

本システムの動作を説明する。まず、内容決定部では発話内容 (what-to-say) を決定し、発話ゴールとして発話プラン生成部に渡す。ただし、内容決定部は現在未実装のため本研究では発話ゴールを人手

```
{refer(Nai),
 type(Nai, '@ない'),
 type(Yoyuu, '@余裕'),
 attrib(Nai, theme, Yoyuu)}
```

図 1: 発話ゴールの例

で作成することによって内容決定部の代用としている。生成する対象領域によって、発話ゴールの種類は変更される。次に、発話プラン生成部は、発話ゴールから発話プランの断片を作成し、表層化できる構造の発話プランに組み上げて表層生成部に渡す。最後に、発話プランを受け取った表層生成部が、表層の制約上発話可能な構造を選択し、日本語を生成する。

発話プランは型付き素性構造で表現されており、発話プラン生成部と表層生成部は単一化を用いて実現されている。この単一化を行うために、SICStus prolog 上で動作する IMS, Stuttgart の Comprehensive Unification Formalism (CUF) 型付き単一化システム [7] を用いている。

本研究では、以下の 3 種類の発話ゴールを使用する。

- refer(inst1): 個体 inst1 を参照する
- type(inst1, type1): 個体 inst1 の型を type1 と指定する
- attrib(inst1, attr1, inst2): 個体 inst1 の属性 attr1 の値は個体 inst2 と指定する

例として図 3 の発話プランを生成するような発話ゴールを図 1 に示す。

発話プランは発話プラン生成部から表層生成部に渡される構造で、型付き素性構造である。その一例を図 3 示す。ここで、 '@余裕' などのように '@' で始まる項は型を表し、 '+時' などのように '+' で始まる項は素性を表している。

発話プラン生成部には属性 attr1 を素性 sem に変換する知識が与えられている。発話プラン生成文はこれを用いてはじめに図 1 の発話ゴールすべてを素

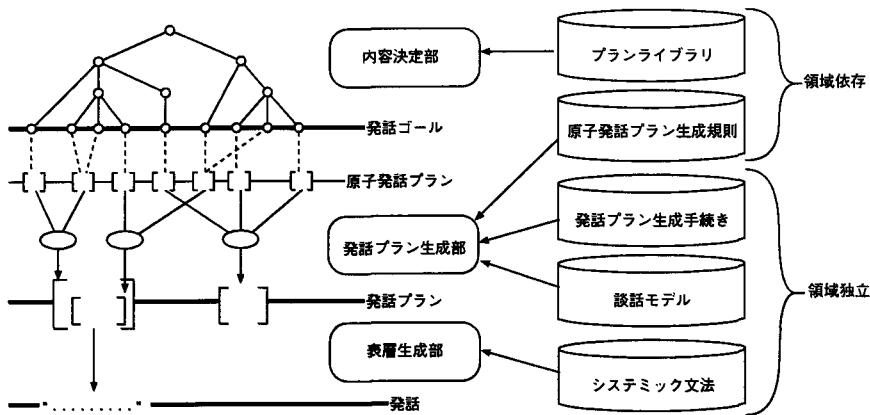


図 2: システム構成

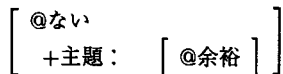


図 3: 発話プランの例

性構造に変換し、次に単一化を用いて図 3 のような発話プランに組み上げる。

表層生成部は発話プランから日本語文を生成するために以下の知識を持つ。

- 単語辞書: 型を単語に変換する
- 選択制約: 単語の意味に関する共起制約
- 順序制約: 単語の順序に関する共起制約

これらは CUF 上にシステミック文法を用いて体系化された言語知識として実装されている。

3 省略表現生成モジュール

3.1 省略表現に関する調査

日本語で省略される可能性がある要素は補語、主題、述語、助詞、主節などである。省略表現生成モジュールを作成するにあたり、省略・照応表現を分析するために日本経済新聞の社説 20 日分を人間の視

察により調査した。この調査から次の知見を得た。

省略された要素

調査対象とした社説において省略されていた要素は以下のようなものであった。

- 述語の格要素全体
- 名詞句内でヘッドの名詞を修飾している名詞句+格助詞
- 名詞句内でヘッドを修飾している名詞句を残してヘッドの名詞が省略される

述語の省略や代用動詞化の事例はなかった。また、ほとんどすべての省略表現が前方照応であった。

省略された格要素

藤澤ら [4] の調査では、省略句の大半は主格の後置詞句であるという結果が出ている。今回の調査でも同様に主格の省略が多かった。特に先行詞が文を跨いでいる場合には主格の省略がほとんどであった。先行詞が文内にある場合は、従属節と主節で共通の要素のうち語順が後ろの要素が省略される場合が多かった。

照応表現については調査対象とした例文が少な

ぎ、一般的な知見を得ることはできなかった。

3.2 省略表現の生成手法

本研究では、文間照応による省略を復元した発話ゴールを人手で作成し、省略表現生成モジュールを組み込んだ日本語文生成システムにこの発話ゴールを与え、省略を含む日本語文を生成する。

省略表現の生成手法としては、文間照応による省略表現は以下のルールで生成することにした。

ルール 前3文以内に主格だった要素が再び主格になった場合、それを省略する。

具体的には、発話プラン生成部の入力である発話ゴールを保存しておき、前3文以内の発話ゴールを参照してこのルールに合致する発話ゴールを削除することで省略する。

文内照応による省略表現の生成手法としては共通の要素のうち語順が後ろの要素を省略することが考えられる。しかし、本システムでは語順と個体を同時に知ることができるモジュールがないため、文内照応による省略表現生成機能の実装を見送った。

4 述部生成機能の拡張

本研究で生成対象とする社説では、石澤の日本語文生成システム [6] では生成できない多くの述部表現が出現するため、本システムの述部生成機能を拡張した。

日本語の述部の構成要素は述部のヘッドである動詞、形容詞、形容動詞以外にも様々な要素が含まれる。日本語の述部を生成するには、入力の指定に従って生成すべき単語を選択しながら述部のすべての単語を日本語として正しい順序に並べなければならない。このとき、活用をする単語については活用語尾の決定もする必要がある。

以上の機能を実現するために、基礎日本語文法 [3]、SFTB [8]、日本語のモダリティ [1]、EDR 日本語単語辞書 [2] を用いて以下の知識を表層生成部に与えた。

- 述部で生成される単語の辞書

先頭	
ヘッド、(動詞、形容詞など ボイス	
ます丁寧表現	認め方
認め方	テンス
テンス	価値判断
価値判断	真偽判断
真偽判断	表現類型
表現類型	です丁寧表現
末尾	

表 1: 述部の語順制約

- 活用型に関する文法知識
- 語順制約 表 1 参照
- 接続制約 EDR 日本語単語辞書の接続表

以上の知識を使って以下の手順で述部を生成する。ここでカテゴリーとは表 1 の「ヘッド」、「認め方」などを指す。

- 1 入力によって生成する単語が一意に決定するカテゴリーはその単語を出力する
- 2 生成する単語が一意に決定しなかったカテゴリーは入力の制約内で生成可能な単語をひとつ出力する
- 3 出力された単語のうち活用する単語については活用語尾を決定する
- 4 出力されたすべての単語を表 1 の順にならべ、EDR 接続属性と接続テーブルを使って接続関係を確認する
- 5 接続関係の確認に成功したら、それを結果として出力する 失敗した場合、2 に戻り別の単語の組み合わせを試す

5 実験

5.1 実験方法

日本経済新聞 1998 年 11 月 12 日の社説 [9]のうち、25 文からなる「消費税は下げられない」について文間照応による省略を復元した発話ゴールを人手

で作成し、生成システムに与えて日本語文を生成させた。

省略されている要素は人手で探した。藤澤ら [4] の手法と同様に、はじめに述語を探す。次に、述語のとり得る格要素を同一文中から探す。同一文中になかった格要素は、それ以前の文の要素からもっとも適切なものを選んだ。

発話ゴールを作成する上で、生成システムの能力不足で生成できない部分はあらかじめ削除した。削除部分は 25 文中に一度しか出現しないため、本実験の目的である省略表現の生成には影響しない。

5.2 実験結果

25 文の発話ゴールに対して、24 文の生成文を得た。原文で文間照応による省略がある文は 25 文中 8 文であり、省略すべき格要素は 9 個であった。生成文ではこのうち 4 文の 5 要素が省略された。原文で省略されていない要素が省略されることはなかった。

6 まとめ

本研究では、日本語文生成システムに省略表現生成機能と新しい表層生成部を組み込み、25 文からなる日本経済新聞の社説の生成実験をおこなった。

省略表現生成機能については、少ないルールを用いて原文に近い文が生成できた。

生成システムの表層生成部では、生成できる述部表現を豊富にするために述部から生成される単語の分類をおこない、EDR 日本語単語辞書の接続表を組み込んだ。生成実験から述部ランクの機能はアスペクト表現を除いて十分な機能があることを確認した。

今後は、まず未実装となった文内照応による省略表現の生成機能を実装することと、照応表現の生成手法の調査および論説文以外の領域への応用などが考えられる。

参考文献

- [1] 仁田義雄, 益岡隆志編, 「日本語のモダリティー」, くろしお出版, 1989
- [2] (株) 日本電子化辞書研究所, EDR 日本語単語辞書, 1993
- [3] 益岡隆志, 田窪行則, 基礎日本語文法, くろしお出版, 1989
- [4] 藤澤, 増山, 内藤, 日本語文章における照応・省略現象の基本的検討, 情報処理学会研究報告, 91-NL-86-6, 1991
- [5] 坂庭克幸, 単一化を利用した漸進的発話生成に関する研究, 修士論文, 東京工業大学, 1997
- [6] 石澤宏明, 粒度の細かい発話ゴールを用いた漸進的発話生成に関する研究, 修士論文, 東京工業大学, 1998
- [7] J.Dörre, G.Dorna and Jörg Junger, *The CUF User's Manual*, Institute für maschinelle Sprachverarbeitung (IMS), Universität Stuttgart, Germany, 1996.
- [8] 佐野 洋, Sftb, Institute for New Generation Computer Technology Sixth Research Laboratory, 1992
- [9] (株) 日本経済新聞社, 日本経済新聞 社説, 1998