

選択可能なスタイルを用いた文書作成支援システム

6Q-7

山崎通弘[†]山村毅[‡]大西昇^{†‡}[†]名古屋大学大学院工学研究科情報工学専攻[‡]愛知県立大学情報科学部[‡]理化学研究所 バイオミメティックコントロール研究センター

1 はじめに

論文などの半定型文書では論旨が明確であることが重要である。しかし作成途中で論旨を見失うことがしばしばある。半定型文書では記述する内容は異なるものの、ある種の決まった論理の流れがある。これを利用することによって、明確な文書の作成が容易になる。このような考え方に基づいた手法として TSML (Text Structure Markup Language) [3] などがある。

しかし、これらの手法で使用される構造などは、ある特定の文書の形式に依存するため、論理の流れは一つに固定され、単調な文書しか作成できない。そこで、本研究ではスタイルという概念を用いることにより、異なる型の文書への対応を行う手法を提案する。ここでスタイルとは、半定型文書の形式を利用して、文書を構造化するのに必要な項目の集合とする。

本稿では、2節で文書作成支援の手法と文書の構造に関して述べる。3節で TSML とその問題点について述べ、4節でスタイルに必要な項目についての考察を行い、5節でシステムのプロトタイプを示す。最後に6節をまとめとする。

2 文書作成支援の手法

2.1 文書作成支援の方向性

文書作成支援には三つの方向性がある。

まず語句や文法に関する支援であり、「推敲」[1]などがすでに市販されている。これらは文法や語句の不適切な部分を検出することができるが、表層的な誤りのみであり、論旨に対する支援は対象にしていない。

次に発想に関する支援であり、KJ法 [2] のように自

由連想とその整理を基本とするものである。しかし、これは「何を書くべきか」への支援であり、論旨の支援、すなわち「どのように書くべきか」への支援は考えられていない。

最後に文章の構成に関する支援であり、TSMLのように、半定型文書の内の技術論文の序論を対象として、論旨を支援するものである。本稿では半定型文書の作成支援を対象とするので、この手法が適当である。

2.2 文書の構造

論旨を見失う原因としては、パラグラフの間の繋がりがおかしくなることや、文章の全体の構成と各部分の詳細の見通しが悪くなることが考えられる。

これらの問題の解決のために、文章の全体を眺める視点と局所を詳しく見る視点の二つの視点から考える。前者の視点からは文章全体の構成を表す構造が、後者からは各文間の繋がりの構造を見つけることができる。前者を文脈構造、後者を修辞構造と呼ぶことにする。

これらを明示することにより、論旨を意識した文書の作成を支援できる。また、各構造の基本単位が接続表現を生成するため、おかしいパラグラフ間の繋がりが無くなる。さらに、全体への視点と局所への視点を分けることにより、末端の記述に注意が引かれて文書の見通しが悪くなることも避けることができる。この考えに基づいたものに、TSMLがある。

3 TSML

3.1 概要

TSMLは、マークアップ方式による文章構造の記述言語であり、役割コマンド、関係コマンド、グループ化コマンド、節指定コマンドの4種類のコマンドから構成される。

役割コマンドは文脈構造を構成する役割を指定する。関係コマンドは修辞構造中の主従関係を指定し、グループ化コマンドはグループ関係を指定する。節指

Writing Assistant system using changeable styles

Michihiro Yamazaki[†], Tsuyoshi Yamamura[‡],

Noboru Ohnishi^{†‡}

[†]Nagoya University

[‡]Aichi Prefectural University

[‡]RIKEN BMC Research Center

定コマンドは、主従関係の順序の入れ替えの際に、共通して先に述べられる節を指定するためのものである。

TSMLによる記述例を以下に示す [図 1]。

```

\Problem{
  \clause{TSMLでは}
  \Parallel{
    \item{
      \reason{構造の数が固定されている}
      異なる文体への拡張性が欠けている
    }
    \item{
      \reason{
        \reason{生成規則が明確に定義されていない}
        実際に使用するときには、
        文献で提示されているシステムに依存している
      }
    }
    \Parallel{
      \item{単調な文書しか生成できない}
      \item{新たな表現形式を追加できない}
    }
  }
}

```

図 1: TSML の記述例

3.2 問題点

TSMLでは対象を論文の序論に限定している。そのため役割コマンドの種類は序論用に制限されている。論文の本論も対象にするためには、コマンドの追加が必要となってくるが、単純に追加する方法では、対象文書を増やすたびにコマンドの追加と実装系の修正が必要となる。

しかし、コマンドを定義する方法を定め、対象文書ごとに必要なコマンドの種類や論理の流れなどを設定することにより、新たな文書を対象にすることが容易になる。このためにスタイルを用いる。

4 スタイル

4.1 スタイルに必要な項目

スタイルを作成するために、必要な項目を文脈構造と修辭構造の視点から考える。

文脈構造は役割コマンドの繋がり構造で表現される。そのため使用する役割コマンド群を定義する必要がある。そして、それらを繋げるためには、役割コマンドの基本配列規則とその配列の入れ換えの制約条件が必要になる。また、役割コマンドと修辭構造を構成する各コマンドは接続表現の生成規則が必要である。

4.2 スタイルの定義

役割コマンドの集合を R 、修辭コマンドの集合を M とすると、スタイルは図 2 のように定義できる。 R_{sub}

はスタイルに使用する役割コマンドの集合である。 O はスタイルの持つ基本序列である。 P は役割同士の位置の制約の集合であり、 $pos(a, b)$ は連続する役割が a と b であることが許されれば 1 で、許されなければ 0 である。 C_R は役割コマンドの生成規則の集合であり、 C_M は修辭コマンドの生成規則の集合である。 $connect(a, b)$ は連続するコマンド a と b を接続する規則である。

$$\begin{aligned}
 S &= (R_{sub}, O, P, C_R, C_M) \\
 R_{sub} &= \{r \mid r \in R\} \\
 O &= \{r^* \mid r \in R_{sub}\} \\
 P &= \{pos(a, b) \mid (a, b) = R \times R\} \\
 C_R &= \{connect(a, b) \mid (a, b) = R \times R\} \\
 C_M &= \{connect(a, b) \mid (a, b) = M \times M\}
 \end{aligned}$$

図 2: スタイルの定義

4.3 スタイルの利点

役割コマンドの種類を変更することにより、異なる型の文書への対応が可能になる。また、各生成規則を変更することにより、複数の文体による文書を作成することができる。

そしてスタイルでは、論理の流れは役割コマンドの序列で表されるため、複数の対象の論理の流れに対応することができる。

5 おわりに

本稿では、半定型文書の作成支援のために選択可能なスタイルを提案し、システムのプロトタイプを作成した。

今後の課題としては、システムの入力インタフェースの改善や入力時の構造解析、構造変更時の代名詞処理などの機能拡張と使い易さの評価法の考察などが挙げられる。

参考文献

- [1] 牛島和夫, 菅沼明, 富士通ソーシアルサイエンスラボラトリー: 推敲マニュアル (入門編, 解説編), 岩波書店 (1993).
- [2] 川喜田二郎: KJ 法, 中央公論社.
- [3] 岩田芳明, 山村毅, 大西昇: “マークアップ方式による文章作成システム”, 電子情報通信学会技術研究報告 NLC97-60.