

文体統一のための階層型文体認識変換方式

2F-4

衣川幸恵, 久保田淳市
松下電器産業(株) 情報通信関西研究所

1. まえがき

近年、文書から誤りや表記の不統一を検出して校正推敲を支援する方式・システムがいくつか提案・実施されている。そのうち特に、待遇表現に関する興味と実用性の高さから文体表現統一方式についての検討が多い。

本稿では、実態に即した文体の分類とそれらの不統一を効果的に解消する階層型文体認識変換方式について報告する。

2. 文例調査

従来、文体分類は「デス・マス調」の敬体と「ダ・デアル調」の常体として扱われることが多かったが実態はより複雑である。[1] A4換算130枚程度のワープロ文書から句点直前文節を抽出し出現頻度を調査した。(表1 表中<その他>とは「…致します」のようにデス・マスを削除、置換するだけでは文体変換が不十分なものである。)

表1 文末表現頻度分布

常体	<…だ>	1%
	<…である>	9%
	<形容詞>	9%
	<補助動詞>	15%
	<動詞>	66%
敬体	<…です>	8%
	<…ます>	25%
	<その他>	67%

敬体のうち、デス・マスの助動詞で文体が規定されるのは全体の1/3程度である。
また、上記の表1には明示的に現れていないが、「めしを食らう」の類の表現や口語的表現には単純にデス・マスの追加削除では前後の単語とのつりあいがとれず文体変換を行った結果、文全体としての調和が乱れる場合がある。

3. 文体の大分類

このようなワープロ文章の実態を踏まえ、文体を待遇の度合に応じて4つに大分類した。無論、ダ・デアル混用調、デアル調などを別に設定することも可能だが、文体表現統一の観点から最小限分類すべきものとしては適当な分類であると考える。

①第一類 (口語調)	【適用文章】文芸文、会話文 【表現例】「めしを食らう」
②第二類 〔常体〕 (だ・である調)	【適用文章】説明文、論説文 【表現例】「政府の責任である」
③第三類 〔敬体〕 (です・ます調)	【適用文章】説明書、マニュアル 【表現例】「右に倒します」
④第四類 〔敬体〕 (ございます調)	【適用文章】手紙文 【表現例】「対処致します」

*(括弧内は待遇の程度を表現するための俗称)

4. 認識変換方式

これらの分類のうち、特に出現頻度の高い二類から四類の文体表現を統一するための基本システムを試作した。

システムは、入力文を形態素解析し、文体を特徴付ける助動詞、補助動詞を検出し文体を認識する。指示文体と異なる場合、待遇表現を除いた標準表現に変換する。最後に、目標とする文体に応じて目的とする文体の表現を生成する。(図1) 以下、同一大分類内での統一、各種の文体定義を容

易に実現するための階層型の文体認識変換方式と言い換えて説明する。

(1) 階層型認識変換方式

文体大分類の下位区分として許容文体と目的文体を設けた。文体の大分類の俗称をダ・デアル調などと呼んでも実態は「…なのだ」「…だ」「…なのである」などと様々なものを含んでいる。これらを出来るだけ広く許容するものを許容文体と呼ぶ。これに対し、文章の文体を統一する立場からは、変換する先の文体は標準的なものに限定するのが妥当である。その文体を目的文体と呼ぶ。(表2に許容文体、目的文体の一部の例を「ダ・デアル調」について左に示す。

文体認識する際は許容文体の範囲で判定し、認識結果が目的文体でない場合に限って目的文体に変換することで文体統一機能が実現できる。(図1)
この、許容文体、目的文体の設定を制御により、各種の文体大分類を実現することができる。また、同一大分類内の文体を標準の目的文体表現に統一することも可能となる。

ダ・デアル調	
許容文体	目的文体
美しい	美しい
美しいのだ	
美しいのである	
書く	書く
書くのだ	
書くのである	
本だ	本だ
本である	本である
本なのである	
本なのだ	

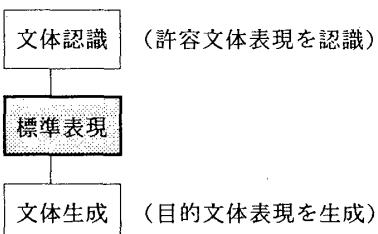


図1 認識変換処理手順

(2) 言い換え

文体変換は、助動詞、補助動詞の置換等で対応するものと語彙の言い換えで実現されるものがある。言い換え効果の高い第一類・第四類の動詞について試行的に語彙抽出し、可能な言い換え語を対応付けた。

第1類動詞 【例】のさばる、こましゃくれる

第4類動詞 【例】拝受する、仰せつかる

5. おわりに

基本システムにて例文を使った詳細評価をすすめている。あわせてシステムの拡張を検討している。

[参考資料] [1]図説日本語 常体と敬体-3 p379 林大