

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られる。

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