

# 印象尺度に基づいた文章言い換え手法

吉岡直輝<sup>†</sup>松本 和幸<sup>‡</sup>北 研二<sup>‡</sup>任 福継<sup>‡</sup><sup>†</sup> 徳島大学大学院 先端技術科学教育部<sup>‡</sup> 徳島大学大学院 ソシオテクノサイエンス研究部<sup>†</sup>yoshioka-naoki@iss.tokushima-u.ac.jp<sup>‡</sup>{matumoto, kita, ren}@is.tokushima-u.ac.jp

## 1 はじめに

本稿では文中の単語を、同じ意味の、別の印象を持つ単語に置き換えることで、印象の違った文に変換する手法を提案する。友人や家族とコミュニケーションをとる際に用いるような単語の印象を、Casual (柔かい表現)、目上の人とコミュニケーションをとる際に用いるような単語の印象を Formal (堅い表現) の尺度と定義し、新聞記事とブログを基に、単語の Casual/Formal の印象付与を自動で行う。新聞記事は短い文で的確にニュースのポイントを読者に伝えるため、会話では使用されることが少ない漢字を多く用いる。このことから、新聞記事で多く用いられる単語を Formal な単語として分類する。ブログは個人的な体験や日記を掲載し、臨場感を出すためにくだけた会話表現を多く用いる傾向がある。このことから、本実験では、ブログで多く用いられる単語は Casual な単語として分類する。

## 2 従来研究

鍛冶ら [1] の研究では、用言を書き言葉から話し言葉に変換する手法が提案されている。しかし、名詞については Casual な表現から Formal な言葉への変換は行っていない。また、田添ら [2] の研究では、動詞を敬語表現に言い換える手法が提案されている。しかし、俗語表現や古風な表現への変換には対応していない。前者の場合 (a) から (b) のような言い換えになり、後者の場合 (a) から (c) のような言い換えになる。入力文が (a) のような場合、動詞部分を言い換えても文全体の印象は変えることは難しい。入力文が (a) の場合、(d) のように名詞部分も言い換えるべきである。そこで、動詞に加え名詞も変換候補とした。

(a) ギャル と ランチ を 食べた。(b) ギャル と ランチ を 食事した。(c) ギャル と ランチ を 頂いた。(d) 女性 と 昼食 を 頂いた。

本研究では、言葉のニュアンスの違いについて書かれた語感の辞典 [3] を参考に、Formal な表現を「学術的な表現」「式典などの改まった表現」「古風な表現」とした。また、Casual な表現を「くだけた会話で用いる表現」「俗語表現」と定義し、単語を Casual/Formal に分類した印象辞書を構築する。

## 3 印象辞書

印象辞書とは、単語ごとに印象の属性を割り振られた単語をまとめたものである。

本研究では、新聞記事コーパスとブログ記事コーパスを用いて基本的な単語の Casual/Formal を表 1 に示すように自動分類し、単語印象辞書を構築した。新聞記事コーパスは毎日新聞コーパス 869,423 記事を用い、ブログコーパスは Web からランダムに収集した 66,386 文を用いた。記事ごとの単語の出現頻度を算出し、上位 200 までの単語を対象とし、ブログのみに出現した単語を Casual な単語とし、新聞記事のみに出現した単語を Formal な単語とし、両方に出現したものを Normal な単語と分類した。

表 1: 単語の Casual/Formal 分類

Casual	ブログ記事のみ
Formal	新聞記事のみ
Normal	両方に出現

## 4 提案手法

本手法は印象尺度に基づいた文章言い換えを行う。以下で本手法の手順を示す。通常の文を入力する。入力文を ChaSen[4] を用いて形態素解析する。形態素ごとに変換可能語の識別をする。変換可能語の条件は名詞、動詞かつ単語印象辞書に含まれるとする。変換可能語の名詞部は文脈類似語データベース[5]を用い、単語印象辞書内から同義語を収集する。また、動詞部は分類語彙表と日本語 WordNet を用いて収集する。その後、単語印象辞書を用いて Casual/Formal の一方の単語に絞り込み、組み合わせで文を出力する。本手法のシステムの流れを図 1 に示す。

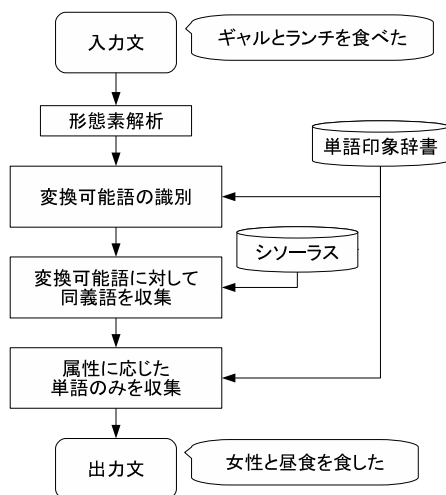


図 1: システムの流れ

## 5 実験

3 節で述べた手法で分類した Formal な単語 144 語と Casual な単語 242 語を被験者へのアンケート調査により印象付けの精度を検証した。アンケートの回答と本手法の分類結果を比較し一致した場合を正解とし正解率を算出した。その結果を表 2 に示す。また、前述の 386 単語と語感の辞典を組み合わせで単語印象辞書を構築した。本手法で分類した単語と語感の辞典で分類した単語が重複した場合は、語感の辞典で分類した印象を優先し、語感の辞典に掲載されていない 138 単語を追加した。構築した単語印象辞書を用いて 30 記事のブログに対して言い換え実験を行った。変換した単語に対して正解率を算出した。その結果を表 3 に示す。

表 2: 実験結果 1

	Formal	Casual	合計
正解率	89.3	66.0	80.6

表 3: 実験結果 2

	Formal	Casual	合計
正解率	49.4	33.5	41.5

## 6 考察

実験 1 の結果、出力が Casual でアンケートが Formal のものが「はう」「過疎」などがあり、複数の意味を持つ単語が誤った。また、出力が Formal でアンケートが Casual のものが「会長」「意味」などがあり、会話でもしばしば用いられる表現だが、ブログで頻出しないため誤ったと考えられる。

実験 2 の結果、「その上」の「上」ような非自立語や「大変そう」の「そう」のような接尾語は誤り率が高かった。今後これらを前後の名詞と連結するか変換候補から除外する処理を行う予定である。

## 7 おわりに

本稿では、印象尺度に基づいた文章言い換え手法について述べた。本手法により、語感の辞典とブログコーパス、新聞記事コーパスを用いることで単語印象辞書を作成した。また、作成した単語印象辞書を用いて印象に応じた言い換えを行うことができた。

## 謝辞

本研究の一部は、科学研究費補助金・挑戦的萌芽研究（課題番号：21650030）により行われた。

## 参考文献

- [1] 鍛冶 伸裕, 岡本 雅史, 黒橋 禎夫: “ WWW を用いた書き言葉特有語彙から話し言葉語彙への用言の言い換え ”, Journal of natural language processing 11(5), 19-37, 2004-10-10
- [2] 田添 丈博, 渡辺 千亜季, 椎野 努: “ 敬語表現への言い換えに関するコンピュータモデルの構築 (言い換えと翻訳) ” 情報処理学会研究報告. 自然言語処理研究会報告 2005(94), 1-6, 2005-09-29
- [3] 中村 明: “ 語感の辞典 ”, 東京堂出版, 2010
- [4] 松本 裕治: “ 形態素解析システム「茶釜」 ”, 情報処理学会誌 Vol.41 No.11 p1208-p1214(2000)
- [5] 風間 淳一, Stijn De Saeger, 鳥澤 健太郎, 村田 真樹: “ 係り受けの確率的クラスタリングを用いた大規模類似語リストの作成 ”, 言語処理学会第 15 回年次大会, pp. 84-87, March, 2009 .