

感情語辞書を用いた中国小説の感情分析

Emotional Analysis of Chinese Novels Using an Emotional Dictionary

黄永輝[†] 上原 稔[†] 安達 由洋[‡]
Huang Yonghui Minoru Uehara Yoshihiro Adachi

1. はじめに

近年、中国と日本の文化交流が盛んになり、日本の小説やドラマなどの作品が大量に中国語に翻訳されて中国に輸入されている。我々は、感情語辞書を用いた日本語感情表現分析システム Emotional Expression Analysis System (EEAS) [1]を開発し、EEAS を用いた文学研究のデジタルアプローチ[2]を提案した。また、中国語感情語辞書を作成し、これに基づいた中国語感情表現分析システム Chinese EEAS (CEEAS) [3]を開発している。

本研究では、CEEAS を用いて中国小説の感情分析を行い、感情出現の特徴を調査した。また、日本小説の感情出現との比較を行った。更に、日本小説とその中国語翻訳版を各々EEAS と CEEAS を用いて感情分析し、2 国間での小説の感情表現の差異を調査した。

2. EEAS と CEEAS

我々は感情語辞書を用いて高速に精度よく日本語文の感情を分析する技術の研究を行ってきた。

EEAS は、[喜, 怒, 哀, 怖, 恥, 好, 厭, 昂, 安, 驚, 望] の 11 感情カテゴリに分類した 7,249 語の語彙を持つ感情語辞書を用いており、日本語 10,000 文の 11 次元感情カテゴリベクトルを約 0.92 秒で計算する[4]。

CEEAS は、EEAS と同じ 11 感情カテゴリ [乐, 怒, 哀, 怖, 耻, 喜, 厌, 激, 安, 惊, 望] に分類した 13,142 語の語彙を持つ中国語感情語辞書を用いており、中国語 10,000 文の 11 次元感情カテゴリベクトルを約 1.5 秒で計算する[5]。

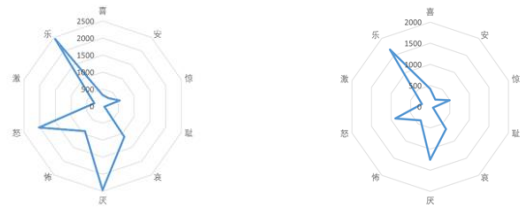
3. 分析対象作品

本研究では、中国小説として莫言の 3 タイトル、韓寒の 3 タイトル、賈平凹の 3 タイトル、余華の 2 タイトルの合計 11 作品を分析対象とした。また、日本語小説として村上春樹の 5 タイトル、池井戸潤の 4 タイトル、東野圭吾の 4 タイトルの合計 13 作品を分析対象とした。

4. 中国小説の分析

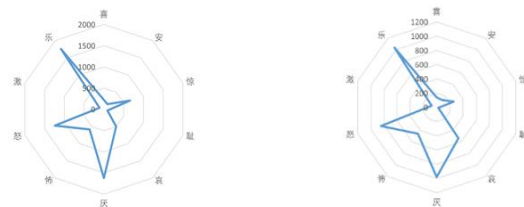
4.1 10 感情レーダーチャート

中国小説の作家毎の感情出現頻度を‘望’感情を除いた 10 感情レーダーチャートで図 1 に可視化した。中国小説の感情語の出現には全作家の作品に共通する大きな特徴がある。すなわち、‘喜’ (中国語の‘乐’)、‘厭’ (中国語の‘厌’)、そして特に‘怒’の感情カテゴリに属する語の出現頻度が高い。また、ネガティブ感情カテゴリに属する‘怒’と‘厭’ (中国語の‘厌’) は出現頻度がほとんど同じである。



(a) 莫言の全小説

(b) 韓寒の全小説



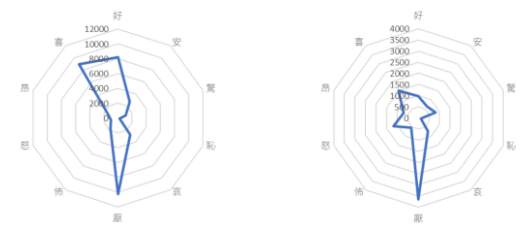
(c) 賈平凹の全小説

(d) 余華の全小説

図 1: 中国語小説の感情レーダーチャート

4.2 日本小説の感情レーダーチャート

対照例として、村上春樹と池井戸潤の全小説の感情語出現頻度の 10 感情レーダーチャートを図 2 に示す。



(a) 村上春樹の全小説

(b) 池井戸潤の全小説

図 2: 日本小説の感情レーダーチャート

5. 日本語から中国語への翻訳小説の分析

5.1 翻訳小説の感情語出現頻度の変化

東野圭吾、池井戸潤、村上春樹の作品の日本語版と中国語翻訳版の感情出現頻度と総文字数の変化を表 1 に示す。表 1 から、東野圭吾のマスカレード・ホテル以外の小説は中国語翻訳版では全体的に少し感情語が増加している。また、各作品の総文字数は中国語翻訳版では元の日本語版より相当減少している。

5.2 翻訳による感情レーダーチャートの変化

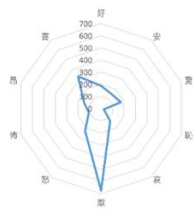
池井戸潤の作品「ロスジェネの逆襲」に出現する感情語出現頻度の感情レーダーチャートを図 3 に示す。図 3(a)は日本語元版を EEAS で分析したもの、図 3(b)は中国語翻訳

[†] 東洋大学 総合情報学研究科 Graduate School of Information Sciences and Arts, Toyo University

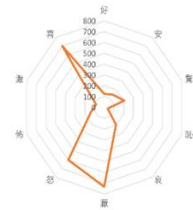
[‡] 東洋大学 工業技術研究所 Research Institute of Industrial Technology, Toyo University

表 1: 翻訳による感情語出現頻度の変化

| | 東野圭吾 | | 池井戸潤 | | 村上春樹 | |
|-------------|------------|--------|----------|---------|------------|------------|
| | マスクレード・ホテル | 探偵ガリレオ | ロスジェネの逆襲 | 銀翼のイカロス | 騎士団長殺し 第1部 | 騎士団長殺し 第2部 |
| 感情出現頻度 (CN) | 30% | 33% | 54% | 56% | 38% | 34% |
| 感情出現頻度 (JP) | 31% | 31% | 48% | 47% | 33% | 30% |
| 総字数 (CN) | 177547 | 103127 | 159565 | 165455 | 195561 | 204575 |
| 総字数 (JP) | 247085 | 141540 | 188591 | 188396 | 307668 | 336033 |



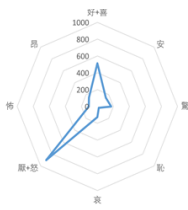
(a) 日本語元版



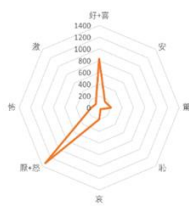
(b) 中国語翻訳版

図 3: 「ロスジェネの逆襲」の感情レーダーチャート

版を CEEAS で分析したものである。図 3 が示すように、「ロスジェネの逆襲」の中国語翻訳版は日本語元版と比較すると、「喜」(‘乐’)と「怒」の感情カテゴリ出現頻度が顕著に多くなっている。しかし、「ロスジェネの逆襲」の日本語元版と中国語翻訳版の感情カテゴリ出現頻度を、「喜」と「好」をまとめ、「怒」と「厭」をまとめた修正 8 感情レーダーチャートで可視化すると図 4 になり、ほとんど同一となる。



(a) 日本語元版



(b) 中国語翻訳版

図 4: 「ロスジェネの逆襲」の修正感情レーダーチャート

5.3 感情語辞書の特徴

感情語辞書を用いた感情分析では精度が、辞書の特性・精度に大きく依存する。EEAS で用いる日本語感情語辞書は日本人 5 名の多数決を取って各感情語が属する感情カテゴリを決定している。また、CEEAS で用いる中国語感情語辞書は中国人 7 名の多数決で各感情語が属する感情カテゴリを決定している。図 4 に示した結果を見ると、各辞書で「喜」カテゴリと「好」カテゴリ、および「怒」カテゴリと「厭」カテゴリに登録している感情語の再検討も課題である。表 2-1 と表 2-2 に日本語感情語辞書 Emotional Word Dictionary (EWD) と中国語感情語辞書 Chinese EWD (CEWD) に収録している 11 感情カテゴリ別の感情語数を示す。

表 2-1: 各感情カテゴリに収録している感情語数 (前半)

| | 喜 | 怒 | 哀 | 怖 | 恥 | 好 |
|------|------|------|------|-----|-----|-----|
| EWD | 1061 | 631 | 735 | 485 | 222 | 965 |
| CEWD | 2382 | 6163 | 1781 | 945 | 99 | 408 |

表 2-2: 各感情カテゴリに収録している感情語数 (後半)

| | 厭 | 昂 | 安 | 驚 | 望 |
|------|------|-----|-----|-----|-----|
| EWD | 2008 | 730 | 337 | 338 | 370 |
| CEWD | 6851 | 106 | 313 | 239 | 109 |

なお、複数の感情カテゴリに属する感情語もあるので、表 2-1、2-2 の感情語の総和は各辞書の語彙数を超えている。

6. まとめ

CEEAS を用いて中国小説の感情分析を行った。中国 4 作家の作品の感情カテゴリ出現頻度はよく似ており、日本作家とは異なる特徴を持っている。また、日本語小説の感情カテゴリ出現頻度を EEAS を用いて分析し、その中国語翻訳版の感情カテゴリ出現頻度を CEEAS を用いて分析することで、翻訳により感情出現頻度が変化するかどうかを調査した。その結果、「喜」(‘乐’)と「好」(‘喜’)を一つのカテゴリとしてまとめ、さらに「怒」(‘怒’)と「厭」(‘厌’)を一つのカテゴリとしてまとめると、日本語元版と中国語翻訳版の感情カテゴリ出現頻度がほとんど同一になることが分かった。

なお、情報処理技術を用いた文学研究として品詞と意味に基づくクラスター分析を行った論文[6]や、AI 技術による小説の分析サービスなどが報告されている[7]が、本研究のような感情に基づく中国文学分析に関する研究報告は見当たらない。

今後の課題として、様々なジャンルの多くの中国作家の作品を感情分析してそれらの特徴を明らかにすることが挙げられる。また、より多くの日本作家の作品の中国語翻訳版との感情分析・比較をするとともに、逆に中国作家の作品の日本語翻訳版との感情分析・比較も課題である。さらに、これらの分析過程で得られた感情語辞書に関する情報を用いて感情語辞書をアップデートすることも重要な課題である。

参考文献

- [1] 安達由洋, 近藤友啓, 小林孝充, 恵谷菜央, 石井解人, “感情語辞書を用いた日本語文の感情分析”, 可視化情報 Vol.41 No.161 (2021).
- [2] 瀬山透矢, 加藤陸斗, Astremo Amilcare, 天野樹, 中山佳大, 安達由洋, “集合知に基づく現代日本文学研究のアプローチ”, 情報処理学会第 84 回全国大会 (2022).
- [3] 黄永輝, 楊海洋, 冉升, 卜廷君, 李星銳, 成思遠, 趙鑫, 安達由洋, “中国語感情語辞書と中国語感情表現分析システム”, 情報処理学会第 84 回全国大会 (2022).
- [4] 瀬山透矢, Amilcare Astremo, 安達由洋, “ソーシャルメディアおよび EC サイトでのレビュー分析のための‘望’感情の抽出”, FIT2021 (2021).
- [5] Huang Yonghui, Kenshin Tsumuraya, Minoru Uehara, Yoshihiro Adachi, “High-Speed Emotional Analysis of Free-Descriptive Sentences Using an Emotional-Word Dictionary”, IEEE ICCSE2022 (August 2022, accepted).
- [6] 工藤彰, 村井源, 往住彰文, “計量分析による村上春樹文学の語彙構成と歴史の変遷”, 情報知識学会誌 2010 Vol.20, No.2, pp.135-140 (2010).
- [7] ツギクル事務局, “ツギクル”, <https://www.tugikuru.jp/> (2022/06/22 アクセス).