

科学研究費助成事業 研究成果報告書

平成 26 年 4 月 24 日現在

機関番号：32663

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700288

研究課題名(和文) 計算文体論による多種メディアテキスト解析

研究課題名(英文) Analyzing various media texts using computational stylistics methods

研究代表者

鈴木 崇史 (SUZUKI, Takafumi)

東洋大学・社会学部・准教授

研究者番号：70507037

交付決定額(研究期間全体)：(直接経費) 3,000,000円、(間接経費) 900,000円

研究成果の概要(和文)： Webの発達は、既存のメディアに変革を迫りつつある。とりわけ、ブログや掲示板をはじめとするCGM(消費者生成メディア)の発達は、新しいコミュニケーション・スタイルを生みだしている。本研究では、多種メディアテキストを収集し、計算文体論アプローチを用いてこれを分析することで、各メディアの特徴を分析し、コミュニケーション・スタイルの特徴を考察した。同時に、有用な分析手法の検討を行った。

研究成果の概要(英文)： Through the development of the Web, various new media have appeared. In particular, texts in social media such as Wiki, blogs, and SNS are produced by the users themselves, and reveal new styles of on-line communication. This study, by using computational stylistics methods, analyzed the character of ones' communication styles from the various media texts. As well, we developed the effective methods for the analysis.

研究分野：総合領域

科研費の分科・細目：情報学 図書館情報学・人文社会情報学

キーワード：テキスト分析 人文情報学 図書館情報学 計算文体論 計量情報学 デジタル人文学 メディア コミュニケーション

## 1. 研究開始当初の背景

Web の発達 は、既存のメディアに 変革を 迫りつつある。とりわけ、ブログや 掲示板をはじめとする CGM (消費 者生成メディア) の 発達は、新しい コミュニケーション・スタイルを生みだし、新聞、雑誌、書籍をはじめとする書誌メディアから、テレビ、映画など、映像メディアまで、既存のメディアは、この新たなメディアと人々の コミュニケーション・スタイルの変化に対応を 迫られている。すでに、この点をめぐる理論的考察は多くあるものの、今後さらにこれを 実証的研究と 接続することが求められている。

一方、データベースの蓄積と自然言語処理技術の発達により、計算機ベースのテキスト解析は、その射程をますます広げ、大量・多種のテキストデータを新たな 応用へと 役立てることが、学術的、社会的に 急務の課題となっている。とりわけ、テキストから、流行主題や重要トピックのみならず、コミュニケーションの在り方を抽出・分析することへの期待が高まっている。テキストデータに注目し、これを言語学的、社会的に 分析することで、新たなメディアと人々の コミュニケーション・スタイルの変化を 実証的に 明らかにすることができる。

研究代表者は、従来から、大量かつ多種の情報資源の蓄積の中で、テキストによる コミュニケーションに注目し、書き手 (話し手) の態度、価値観、感情など、繊細な情報を抽出しようと試みてきた。とくに、テキストがもつ文体的特徴が、これをあらわすと 考え、これまで、著者推定やジャンル判別などに利用されてきた計算文体論の分析手法を 発展させることで、この点を 明らかにすることができる と考えてきた。この点についてすでに、一部成果を得ているものの、多様な Web テキスト、また、書誌メディア、映像メディアを含む、様々なメディアテキストを 網羅的に 収集し、包括的にメディアの特性と人々の コミュニケーション・スタイルの特徴を 明らかにするには 至っていない。

## 2. 研究の目的

以上のような背景のもと、本研究では、計算文体論アプローチを用いて、多種メディアテキストから コミュニケーション・スタイルを抽出、分析する。本研究では、大量かつ多種のメディアからテキストデータを 収集、電子化し、これを、計算文体論アプローチを用いて分析することで、多様なメディアにおける人々の コミュニケーションの在り方を 実証的に 明らかにすることを 目的とする。これにより、多様なメディアにおける人々の コミュニケーション・スタイルの特徴が 実証的に 明らかになり、将来のメディアの在り方、オンラインでの円滑な コミュニケーションの在り方、書誌メディア、映像メディアの将来、以上を 考え、設計していくために 有益な 実証的知見を得ることを 目的とする。同時に、多

様なメディアテキストを検討することは、言語学分野でのジャンル、レジスター研究、自然言語処理の分野での domain adaptation の問題に対しても、有用なケース・スタディを提供する。

## 3. 研究の方法

本研究では、まず、Blog、Twitter、Yahoo! 知恵袋をはじめとする Web メディア、小説、新聞、雑誌、書籍、音楽 (歌詞)、映像作品 (テレビ番組、映画スクリプト) など、書誌、映像メディアから、テキストデータを 収集、作成し、特徴量を 計算する。次に、多変量解析を適用することで、探索的にテキスト特徴量に影響を与える要因を 考察する。統制要因を 排除した後、機械学習の適用によって、各類型をあらわすテキスト特徴量を 明らかにする。以上をもとに、各メディアの コミュニケーション・スタイルの特徴、位置関係を 明らかにする。初年度は、テキストデータの整備、テキスト特徴量の計算にあて、次年度に、探索的分析、機械学習を用いた特徴分析、最終年度に、メディアの位置関係の分析、研究全体の 総括を行う。

## 4. 研究成果

研究方法に記載の通り、多種データを クロールおよび人手により 収集、電子化するとともに、既存のデータセットを 入手、利用した。同時に、これら 入手したデータについてテキスト特徴量を抽出し、多変量解析、機械学習諸手法を用いて、テキストの特徴分析を行い、言語学的、社会的知見を得た。これをもとに、各メディアの コミュニケーション・スタイルの特徴、位置関係について 考察した。加えて、独自の特徴量を 考案し、また、他のデータをも利用して、分析手法の有効性を 検証することで 方法論的知見を得た。詳細は、以下の 発表論文等に 記載の通りである。今後、さらに、本研究の成果を コミュニケーション指標の作成、検索システムへと 応用することが 課題である。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には 下線)

[雑誌論文] (計 6 件)

SUZUKI, Takafumi, HOSOYA, Mai. Computational stylistic analysis of popular songs of Japanese female singer-songwriters, *Digital Humanities Quarterly*, (accepted) [査読あり].

SUZUKI, Takafumi, KAWAMURA, Shuntaro and AIZAWA, Akiko. Stylistic analysis of text submission to Japanese Q&A communities, *Journal of Quantitative Linguistics*, 19(4), 262-280, 2012 [査読あり].

SUZUKI, Takafumi, KAWAMURA, Shuntaro, YOSHIKANE, Fuyuki, KAGEURA, Kyo, AIZAWA, Akiko. Co-occurrence-based indicators for authorship analysis, *Literary & Linguistic Computing*, 27(2), 197-214, 2012 [査読あり].

SUZUKI, Takafumi. Investigating macroscopic transitions in Japanese foreign policy using quantitative text analysis, *International Relations of the Asia-Pacific*, 11(3), 460-491, 2011 [査読あり].

鈴木崇史・影浦峽. 名詞の分布特徴量を用いた政治テキスト分析, *行動計量学*, 38(1), 83-92, 2011 [査読あり].

ARAKAWA, Yui, KAMEDA, Akihiro, AIZAWA, Akiko, SUZUKI, Takafumi. Adding Twitter-specific features to stylistic features for classifying tweets by user type and number of retweets, *Journal of the American Society for Information Science and Technology*, (accepted) [査読あり].

[学会発表](計 14 件)

SUZUKI, Takafumi, YAMASHITA, Natsumi. Analysis of perspectives in contemporary Japanese novels using computational stylistic methods, *Digital Humanities 2014*, Lausanne (Lausanne University, July 8-11 (accepted)).

SUZUKI, Takafumi, KANOU, Erina, ARAKAWA, Yui. Comparative analyses of textual contents and styles of five major Japanese newspapers, *Proceedings of PACLIC27: the 27th Pacific Asia Conference on Language, Information and Computation*, 2013, Taipei (National Chengchi University, Nov. 22-24), .

SUZUKI, Takafumi, UCHIYAMA, Kiyoko, TOMISAKA, Ryota, AIZAWA, Akiko. Analyzing the characteristics of academic paper categories by using an index of representativeness, *Proceedings of PACLIC25: the 25th Pacific Asia Conference on Language, Information and Computation*, 587-596, 2011, Singapore (Nanyang Technological University, Dec. 16-18).

SUZUKI, Takafumi, HASEGAWA, Shin, HAMAMOTO, Takayuki, AIZAWA, Akiko. Document recommendation using data compression, *Procedia - Social and Behavioral Sciences Volume, Computational Linguistics and Related Fields*, 27,

150-159 (Proceedings of PACLING 2011), 2011, Kuala Lumpur (International Islamic University Malaysia, July 19-21).

ARAKAWA, Yui, YOSHIMOTO, Ryosuke, YOSHIKANE, Fuyuki, SUZUKI Takafumi. An investigation on Twitter use of researchers: user type classification and text analysis, *JADH2013 & DH-JAC2013*. Kyoto (Ritsumeikan University, Sep. 19-21).

ARAKAWA, Yui, KAMEDA, Akihiro, AIZAWA, Akiko, SUZUKI, Takafumi. Stylistic analysis of tweets that are likely to be shared, *JADH2012*, 2012, Tokyo (University of Tokyo, Sep. 15-17).

鈴木崇史・田之上貴之. 複数の中心性指標と中心化傾向を用いた Q&A コミュニティネットワークのカテゴリの特徴分析, 第 9 回ネットワーク生態学シンポジウム, 2012, 宜野湾 (沖縄国際大学, 12 月 15 日-16 日).

鈴木崇史・河村俊太郎・荒川唯・吉元涼介, ランダムフォレストを用いた世論調査データ解析, 日本行動計量学会第 40 回大会抄録集, 187-190, 2012, 新潟 (新潟県立大学, 9 月 13 日-16 日).

狩野恵里奈・荒川唯・鈴木崇史. 内容と文体による五大全国紙の比較分析, 言語処理学会第 19 回年次大会発表論文集, 2013, 名古屋 (名古屋大学, 3 月 12 日-15 日).

小林佳織・狩野恵里奈・鈴木崇史. 女性グループの歌詞の計量テキスト分析, 言語処理学会第 19 回年次大会発表論文集, 2013, 名古屋 (名古屋大学, 3 月 12 日-15 日).

山下夏実・鈴木崇史. 計量テキスト分析を用いた小説の視点研究: 伊坂幸太郎を例として, 言語処理学会第 19 回年次大会発表論文集, 2013, 名古屋 (名古屋大学, 3 月 12 日-15 日).

荒川唯・鈴木崇史・芳鐘冬樹. 専門情報の抽出に向けた専門家の Twitter 利用調査とアカウント分類実験, 第 60 回日本図書館情報学会研究大会発表要綱, 2012, 東京 (日本大学, 11 月 12 日-13 日).

荒川唯・亀田堯宙・相澤彰子・鈴木崇史. Retweet に着目した広がりやすい Tweet の特徴分析, 情報処理学会第 74 回全国大会発表論文集, 1.617-1.618, 2012, 名古屋 (名古屋工業大学, 3 月 6 日-8 日).

青木宣康・河村俊太郎・鈴木崇史. アメリカ合衆国大統領演説の計量テキスト分析, 情

報処理学会第 74 回全国大会発表論文集，  
2.261-2, 262, 2012, 名古屋（名古屋工業大  
学，3月6日-8日）.

〔図書〕(計1件)

鈴木崇史．政治テキストの計量分析，石田  
基広・金明哲編，コーパスとテキストマイニ  
ング，共立出版，2012．

〔産業財産権〕

出願状況（計0件）

取得状況（計0件）

〔その他〕

<http://researchmap.jp/stratovarius/>

## 6．研究組織

### (1)研究代表者

鈴木 崇史（SUZUKI Takafumi）  
東洋大学・社会学部・准教授  
研究者番号：70507037

### (2)研究協力者

河村 俊太郎（KAWAMURA Shuntaro）  
愛知淑徳大学・人間情報学部・講師

荒川 唯（ARAKAWA Yui）  
筑波大学大学院・図書館情報メディア研究  
科・大学院修士課程

青木 宣康（AOKI Noriyasu）  
東洋大学・社会学部・学部学生

狩野 恵里奈（KANOU Erina）  
東洋大学・社会学部・学部学生

小林 佳織（KOBAYASHI Kaori）  
東洋大学・社会学部・学部学生

山下 夏実（YAMASHITA Natsumi）  
東洋大学・社会学部・学部学生

吉元 涼介（YOSHIMOTO Ryosuke）  
筑波大学大学院・図書館情報メディア研究  
科・大学院修士課程

田之上 貴之（TANOUE Takayuki）  
東洋大学・社会学部・学部学生