

令和 4 年 6 月 9 日現在

機関番号：33917

研究種目：若手研究

研究期間：2018～2021

課題番号：18K12194

研究課題名（和文）言語哲学を中心としたヘイト・スピーチの多面的研究

研究課題名（英文）A Philosophy of Language Approach to Understanding Hate Speech

研究代表者

和泉 悠（Izumi, Yu）

南山大学・人文学部・准教授

研究者番号：10769649

交付決定額（研究期間全体）：（直接経費） 1,700,000円

研究成果の概要（和文）：本研究の目的は、ヘイト・スピーチを中心とする、市民社会において不信と断絶を生む言語活動を、言語哲学とその関連分野の研究手法を用いて分析し、理解を与えることである。最初に、差別的語彙や特徴的な構文など、表現そのものの分析を進めた。得られた知見の一つは、そうした表現は、話者が意識的に把握できない、社会的含意をともなう複雑な意味内容を持っているということである。その後、オンライン上も含め、そうした表現が実際に使用される場面を検討した。特に、SNS上での有害な投稿を吟味し、それらを自動的に検出するような技術の可能性を探った。

研究成果の学術的意義や社会的意義

SNSなどを通じて、数多くの人間が表現活動を行う。そのため、どのような表現が市民社会において有害なのか、そしてそれらはどうして有害なのか、ということをも明らかにする必要がある。この研究を通じて、具体的な表現を分析し、それらがどのような内容を持っており、どのように使用された際有害になるのか、ということも一定程度明らかにした。また、そうした有害な言説を、AIを用いて自動的に検出する技術を作るための準備を実際に行い、どのような課題があるのかを特定した。

研究成果の概要（英文）：The objective of this research project is to analyze harmful linguistic activities, such as hate speech, that can create distrust and division among the citizens of a society, using research methods in philosophy of language and the related fields. We first focused on the meaning of linguistic items themselves, including racial and ethnic slurs and constructions used to insult people. One of the findings was that such expressions have complex and socially significant meanings that the speakers cannot be aware of. We then turned to the actual uses of such expressions. In particular, we examined harmful posts on social media and explored the possibilities of technologies for automatically detecting them.

研究分野：言語哲学

キーワード：言語哲学 意味論 語用論 ヘイト・スピーチ

1. 研究開始当初の背景

ヘイト・スピーチに関するこれまでの研究の多くは、その規制が表現の自由と相反せずに正当化できるのかといった、法学・政治学・社会学的問いを中心として展開されてきた。そのため、そもそもヘイト・スピーチが何であるのか、その言語的側面はどのようなものかという、基礎的な研究が不十分であるという指摘があった。

2. 研究の目的

(1) そこで、本研究は、ヘイト・スピーチを中心とする、市民社会において不信と断絶を生む言語活動を、言語哲学とその関連分野の研究手法を用いて分析し、理解を与えることを目的とした。

(2) たとえば、ヘイト・スピーチは単なる言論ではなく、有害である主張されてきたが、なぜそのようなのか、どのような語彙や構文の使用がどのようにして害を生み出すのか、という言語的仕組みの解明を目指した。

3. 研究の方法

(1) 本研究では、第一に、現代英米哲学の一分野である言語哲学と、理論言語学の領域に含まれる意味論・語用論の知見を、ヘイトスピーチをはじめとした有害な言説に適用した。

(2) 第二に、本研究はオンライン上での有害な投稿にも焦点を当て、自然言語処理分野の知見や手法を利用し、有害な言説の分析を行った。

4. 研究成果

(1) これら複数の方法を組み合わせた結果、多岐にわたる研究成果を得ることができた。まずは、特定の言語表現の検討を進め、それら自体が持つ内容を分析した。たとえば、「総称文とセクシャルハラスメント」という論文では、「クジラは哺乳類だ」「象の鼻は長い」といった、科学的一般化を表すために用いられるいわゆる「総称文」の意味内容を分析し、それらがどうして「男性は泣かない」「女性は料理が得意だ」といった、ステレオタイプや偏見の表明につながるのかという点を明らかにした。それには総称文の二つの特徴が関わる。

「AはB」という総称文の特徴の一つは、どれくらいの割合のAがBなのか、という数量にまつわる含意がいまいだという点である。総称文を使うことにより、発言の責任を回避するような形で、特定の集団に対するコメントをすることができてしまう。「男は～」「女は～」と証拠なく述べたとしても、一人くらいはそれが当てはまる人物がいるかもしれない、発言を否定することは難しい。

もう一つの特徴は、総称文を使って何らかの集団について語るとき、その集団をまとめあげ、他の集団と区別する何らかの根本的な性質、本質のようなものがあることが示唆されるというものである。どのような集団分けを行っても、その集団内には多様な人間が含まれる。しかし、総称文を使うことにより、まるで一枚岩であるかのように語ることになる。

他にも、日本語の「～が！」といった、侮蔑的ニュアンスをともなう構文や、差別的語彙とみなされる名詞に対して、哲学と言語学双方の道具立てをを使いつつ分析を与えた。

(2) 特定の言語表現そのものを分析するだけでなく、そうした表現がどのように使用され、どのような害を引き起こすのか、というメカニズムについての知見も得ることができた。たとえば、「ヘイト・スピーチ 信頼の壊し方」という論文では、ヘイト・スピーチが生じさせる害の一種として、政治哲学でしばしば言及される、安心や信頼といった社会的な基盤が壊されうることを示した。類似的な言語表現の理論を応用することにより、侮蔑表現が規約的に持つ意味内容が、会話参加者が相互に参照する想定を半強制的に変更することによって、差別的世観の押しつけが行われる点を指摘した。

他にも、政治的権力を持った主体が、差別的発言への明示的批判を行わないことが、語用論的にどのような効果を生じさせるのか検討した。言語使用がダイナミックに、共通の理解・世界観を変更していくことを示し、差別的表現の使用を「たかがことば」として軽んじる発想が危険であることが導かれた。

(3) 本研究で得られた重要な知見の一つは次のようなものである。言語表現の意味には複合的な側面があり、話者の一人ひとりが明示的に把握している内容はその一部分でしかない。それは日本語待遇表現（敬語や、「さん」といった敬称など）を適切に使用する話者が、その使用のルールを明示的に把握していないのと同じことである。敬語が特定の社会的関係性が成立した条件でのみ使用可能であるように、差別的単語なども特定の社会的・外在的状態の成立をその使用の条件として持つ。話者がどのような心的な態度を持っていようが、敬語使用が社会的関係性を含意してしまうように、差別的単語の使用は、たとえ話者に「悪意がない」としても（本当にそ

れが事実だったとして)、差別的な社会構造を含意し、その維持に貢献してしまう可能性がある。一般的に意識される差別的表現の「意味」と、理論的に解明される内容との乖離は非常に大きい。差別的表現使用の悪影響などは、前者ではなく、後者にもとづいて評価されなくてはならない。

(4) 上述の理論的理解を踏まえ、SNS 上など、インターネットにおける有害な言語使用についても研究を行い、一定の成果が得られた。データサイエンスや自然言語処理の専門家と協力し、究極的にはヘイト・スピーチ自動検出技術の発展を目指し、ツイッターなど SNS を利用した、日本語ヘイト・スピーチデータセットの構築を進めた。

たとえば、「ソーシャルメディアにおけるヘイトスピーチ検出に向けた日本語データセット構築の試案」という論文では、どのようにデータを収集するのか、収集されたデータを誰がどのようにラベルづけするのか、データを適切にラベルづけする作業のためには、どのようなガイドラインが設定されるべきか、といった課題に一から取り組み、試作のデータセットを構築した。

また、「AI はレイシズムと戦えるのか 自然言語処理分野におけるヘイトスピーチ自動検出研究の現状と課題」という論文では、ヘイト・スピーチ自動検出や、カウンタースピーチの自動生成といった、関連技術が実際どのように作成され、機能するのかという点を、具体例をあげつつ概説し、その技術が抱える実際の・社会的課題を指摘し、検討した。たとえば、「自動」に検出されるといっても、どのような判定基準を与えるのか、そしてその基準を学習するための大量のデータを「人力」でひとつひとつ準備しなければならない。その過程における恣意性をどのように克服するかが課題の一部であることを示した。

(5) 本研究の内容を利用して、『悪い言語哲学入門』という一般書を執筆、刊行した。本書はアウトリーチの側面を持つだけでなく、本課題のような言語の有害性に関する研究が、伝統的言語哲学・言語学研究の中でどのような位置づけであるのか、という点も明らかにした。誹謗中傷や嘘といった言語現象は、決して自然言語使用の例外的事例なのではなく、人間の本性に深く根差しており、人間の言語使用を理解する上で避けて通れないものなのである。

5. 主な発表論文等

〔雑誌論文〕 計9件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 6件）

1. 著者名 和泉悠, 仲宗根勝仁, 朱喜哲, 谷中瞳, 荒井ひろみ	4. 巻 1169
2. 論文標題 AIはレイシズムと戦えるのか 自然言語処理分野におけるヘイトスピーチ自動検出研究の現状と課題	5. 発行年 2021年
3. 雑誌名 思想	6. 最初と最後の頁 88-105
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yu Izumi	4. 巻 16
2. 論文標題 The Alleged “Non-Specificity” in Japanese Nominals with Floating Numeral Quantifier	5. 発行年 2021年
3. 雑誌名 Nanzan Linguistics	6. 最初と最後の頁 19-32
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 荒井ひろみ, 和泉悠, 朱喜哲, 仲宗根勝仁, 谷中瞳	4. 巻 27
2. 論文標題 ソーシャルメディアにおけるヘイトスピーチ検出に向けた 日本語データセット構築の試案	5. 発行年 2021年
3. 雑誌名 言語処理学会 第27回年次大会 発表論文集	6. 最初と最後の頁 466-470
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 和泉悠	4. 巻 50
2. 論文標題 「土人が」の語用論	5. 発行年 2019年
3. 雑誌名 メタフュシカ	6. 最初と最後の頁 63-73
掲載論文のDOI (デジタルオブジェクト識別子) 10.18910/73767	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 和泉悠	4. 巻 34
2. 論文標題 「書評 樽本英樹編『排外主義の国際比較』」	5. 発行年 2019年
3. 雑誌名 社会と倫理	6. 最初と最後の頁 142-146
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 和泉悠	4. 巻 69
2. 論文標題 総称文とセクシャルハラスメント	5. 発行年 2018年
3. 雑誌名 哲学	6. 最初と最後の頁 32-43
掲載論文のDOI (デジタルオブジェクト識別子) 10.11439/philosophy.2018.32	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Yu Izumi, Shun Tsugita, and Masaharu Mizumoto	4. 巻 14
2. 論文標題 Knowing How in Japanese	5. 発行年 2019年
3. 雑誌名 Nanzan Linguistics	6. 最初と最後の頁 9-24
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Yu Izumi and Shintaro Hayashi	4. 巻 -
2. 論文標題 Expressive Small Clauses in Japanese	5. 発行年 2018年
3. 雑誌名 New Frontiers in: JSAI-isAI 2017 Workshops Revised Selected Papers, Lecture Notes in Computer Science/Artificial Intelligence	6. 最初と最後の頁 188-199
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-319-93794-6_13	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 和泉悠	4. 巻 68
2. 論文標題 差別語のなにか悪いのか	5. 発行年 2019年
3. 雑誌名 倫理学年報目録	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計15件 (うち招待講演 7件 / うち国際学会 8件)

1. 発表者名 Yu Izumi, Hiromi Arai, Hitomi Yanaka, Katsuhito Nakasone, Heechul Ju
2. 発表標題 Abusive Tweets in Japanese during the COVID-19 pandemic
3. 学会等名 The 3rd International Workshop on Hate Speech in Asia and Europe: Pandemic, Fear, and Hate (国際学会)
4. 発表年 2021年～2022年

1. 発表者名 和泉悠, 荒井ひろみ, 朱喜哲, 仲宗根勝仁, 谷中瞳
2. 発表標題 哲学の応用と社会実装 ヘイトスピーチをめぐる文理共創研究の可能性と課題
3. 学会等名 応用哲学会第十三回年次研究大会
4. 発表年 2021年～2022年

1. 発表者名 Yu Izumi
2. 発表標題 The Semantics of 'Dojin': A Philosophy of Language Approach to Ethnic Slurs
3. 学会等名 Hate Speech in Asia: Challenges and Solutions (国際学会)
4. 発表年 2020年～2021年

1. 発表者名 和泉悠
2. 発表標題 意味研究の実験的手法
3. 学会等名 科学基礎論学会（招待講演）
4. 発表年 2020年～2021年

1. 発表者名 荒井ひろみ・和泉悠・朱喜哲・仲宗根勝仁・谷中瞳
2. 発表標題 ソーシャルメディアにおけるヘイトスピーチ検出に向けた日本語データセット構築の試案
3. 学会等名 言語処理学会
4. 発表年 2020年～2021年

1. 発表者名 和泉悠
2. 発表標題 「ヘイトスピーチ」の概念工学へむけて
3. 学会等名 日本倫理学会第70回大会
4. 発表年 2019年

1. 発表者名 Yu Izumi and Kent Erickson
2. 発表標題 Polymorphic Predicativism
3. 学会等名 The 27th Japanese and Korean Linguistics (国際学会)
4. 発表年 2019年

1. 発表者名 和泉悠・朱喜哲・仲宗根勝仁
2. 発表標題 ヘイトスピーチ 信頼の壊しかた
3. 学会等名 ワークショップ 差別の倫理学・言語哲学・社会学(招待講演)
4. 発表年 2019年

1. 発表者名 Yu Izumi
2. 発表標題 Polymorphic Predicativism
3. 学会等名 Nanzan Workshop on the Foundational Issues in Linguistics and Philosophy of Language(招待講演)(国際学会)
4. 発表年 2019年

1. 発表者名 和泉悠
2. 発表標題 「土人が」の意味論と語用論
3. 学会等名 応用哲学会第十回年次研究大会
4. 発表年 2018年

1. 発表者名 Yu Izumi
2. 発表標題 Slurs and Antihonorifics in Japanese
3. 学会等名 The Fourth Conference on Contemporary Philosophy in East Asia(国際学会)
4. 発表年 2018年

1. 発表者名 和泉悠
2. 発表標題 差別語の何が悪いのか
3. 学会等名 日本倫理学会第69回大会 (招待講演)
4. 発表年 2018年

1. 発表者名 Yu Izumi
2. 発表標題 Slurs and Antihonorifics in Japanese
3. 学会等名 Generative Perspectives on the Syntax and Acquisition of Japanese, NINJAL Workshop (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 Yu Izumi
2. 発表標題 The Semantics of 'Dojin': A Philosophy of Language Approach to Ethnic Slurs
3. 学会等名 Hate Speech: International Perspectives in and for Asian Region (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 Yu Izumi
2. 発表標題 Comments on "Essentializing Inferences" by Katharine Ritchie
3. 学会等名 Meaning and Reality in Social Context (招待講演) (国際学会)
4. 発表年 2019年

〔図書〕 計6件

1. 著者名 和泉 悠	4. 発行年 2022年
2. 出版社 筑摩書房	5. 総ページ数 256
3. 書名 悪い言語哲学入門	

1. 著者名 鈴木貴之、和泉悠、笠木雅史、太田紘史、鈴木真、唐沢かおり	4. 発行年 2020年
2. 出版社 勁草書房	5. 総ページ数 216
3. 書名 実験哲学入門	

1. 著者名 斎藤衛、高橋大厚、瀧田健介、高橋真彦、村杉恵子、和泉悠	4. 発行年 2020年
2. 出版社 開拓社	5. 総ページ数 320
3. 書名 日本語研究から生成文法理論へ	

1. 著者名 Masaharu Mizumoto, Jonardon Ganeri, Cliff Goddard, Yu Izumi, Shun Tsugita	4. 発行年 2020年
2. 出版社 Routledge	5. 総ページ数 288
3. 書名 Ethno-Epistemology: New Directions for Global Epistemology	

1. 著者名 蝶名林亮, 和泉悠, 他9名	4. 発行年 2019年
2. 出版社 勁草書房	5. 総ページ数 368
3. 書名 メタ倫理学の最前線	

1. 著者名 小山虎, 和泉悠, 他16名	4. 発行年 2018年
2. 出版社 勁草書房	5. 総ページ数 372
3. 書名 信頼を考える	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------