

技術的人工物による「発話」の道德的身分について ——言語行為論の枠組みを用いた検討

東京大学・日本学術振興会・理化学研究所 AIP 水上拓哉

論文の概要

本論文の目的は、対話システムのような技術的人工物の「発話」の道德的身分を言語行為論の枠組みから考察することにより、従来のロボット倫理学が焦点を当ててきた「道德的行為者性 (moral agency)」概念に頼ることなく技術的人工物の発話のもつ道德的重要性を説明する可能性を提示することにある。

近年の機械学習の理論的進展を背景に、自然言語によってユーザと対話する技術的人工物⁽¹⁾の「発話」の質も向上している。たとえば、“sequence to sequence (seq2seq)” という技術を用いた Neural Conversational Model (Vinyals and Le 2015) は、従来のルールベース型のシステムでは実現不可能だった柔軟な発話生成の可能性を示唆している。このような技術的進展を背景に、対話システムやコミュニケーションロボットといった「発話」する技術的人工物は教育や介護、エンターテインメントといった分野での活躍が期待される。

しかしながら、最近ではこういった技術的人工物の発話がもつ社会的・道德的影響力にも注目が集まっている。たとえば 2016 年に Microsoft が Twitter 上で運用したチャットボット “Tay” は、ユーザからの学習により攻撃的・差別的な内容のツイートしたとして話題となった⁽²⁾。この問題を受けて国内においても対話システムが不適切な発話をしないように開発者が留意する必要性が指摘されている (東中 2016)。“Tay” のような技術的人工物は人間と同等の言語能力や心的状態をもっているとは言い難いにもかかわらず、ユーザや社会に対して無視できない道德的影響力をもっている。このような技術的人工物の発話がもつ道德的重要性について、倫理学においてはいかなる説明が可能なのだろうか。

倫理学側においては「ロボット倫理学 (robot ethics)」の名の下、ロボットや (いわゆる) 人工知能の倫理について哲学的検討がなされてきた⁽³⁾。後述するように、ロボット倫理学においてはしばしば、ロボットを道德的議論の領域に入れるためにそれが「道德的行為者 (moral agency)」であるか否かが問われてきた (Coeckelbergh 2009a; Gunkel 2012)。しかしその一方で、技術的人工物の道德的重要性 (moral significance) を行為者性概念なしで説

明しようとする立場も多数存在する（たとえば Johnson 2006; Coeckelbergh 2009b など）。道徳的行為者性概念とその他の道徳的諸概念（道徳的重要性、責任、価値、非難など）の関係性についてはロボット倫理学において取り組まなければならないメタ的問いの一つである。

本論文では、考察の対象を技術的人工物の発話に限定する。その上で、少なくともその道徳的重要性を説明するのに必ずしも道徳的行為者性概念が要請されるとは限らないことを、言語行為論の枠組みを用いることによって指摘する。なぜならば、言語行為論における発話内行為および発話媒介行為という分類を用いることは、従来の道徳的行為者性概念をめぐる混乱に踏み込むことなく技術的人工物の発話の道徳的重要性を分析する道筋を与えるからである。具体的にはまず、言語行為論における言語行為の分類、すなわち発話内行為と発話媒介行為の分類を確認した上で、そのうち発話内行為がもつ道徳的重要性について検討する。次に、発話内行為の適切性条件に着目し、技術的人工物の発話内行為の可能性についての議論がロボット倫理学においてなされてきた道徳的行為者性をめぐる議論と対応関係にあることを指摘する。最後に、オースティンやバトラーなどを参照してヘイトスピーチのような攻撃的発話の発話媒介的な性質を示唆し、それを技術的人工物の文脈で再考する。

発話内行為がもつ道徳的規範

オースティンから始まりサールが体系化した言語行為論は現在、分析哲学における当時のモチベーションに留まらず、言語学や社会学、情報工学においても知見を提供している。実際、本論文で対象としている「対話システム」においても、ユーザから受け取った発話を理解するためにその発話がどのような発話内行為かを分類するような仕組みが実装されることが多い⁽⁴⁾。本節ではまず、この言語行為論の枠組みを用いて発話内行為とそれが伴う規範の関係に着目し、その道徳的重要性を確認する。

言語行為論の開拓者ともいえるオースティンは『言語と行為』のなかで発話という行為の内実を以下のように分類した（Austin 1962）。

- ・発話行為：一定の意味と言語対象を伴って一定の文を発すること
- ・発話内行為：情報伝達・命令・警告など、慣習的な発言の力をもつ行為
- ・発話媒介行為：説得のように何かを言うことによって引き起こされる行為

これらのうち、発話行為と発話内行為の区別についてはサール（1969）のよ

うにその区別を認めずにまとめて「発話内行為」とする立場もある。本論文においても基本的には発話行為を含んだ形で発話内行為という表現を用いることにする。しかし、発話内行為と発話媒介行為についてはサールも区別しており、本論文でもこれらは区別して議論することにする。発話内行為は、その発話をすることで (in saying) 行為遂行となっており、一方で発話媒介行為のほうはその発話によって (by saying) 行為が達成される。つまり、発話内行為は私たちが行う行為そのものであり、発話媒介行為はそれがもたらす結果のことを指す。後に述べるように、この区別は発話の道徳的性質を考える上で重要な区別である。ひとまず以下では上記の分類のうち「発話内行為 (illocutionary act)」に焦点を当て、その道徳的性質について確認していく。

オースティンによれば「約束」、「言明」、「脅迫」、「感謝」などといった発話は発話内行為であるが、それらがもつ機能は実に多様である。オースティンは発話内行為の「型」について、判定宣告型、権限行使型、行為拘束型、態度表明型、言明解説型の 5 つに分類した。もちろんこれは、彼が反駁しようとしていた立場、すなわち真・偽の二分法と価値・事実の二分法の境界線を混乱させるために作られたもので、彼自身その分類には納得していない。一方サールはこの分類において分類の基準がないと批判した上で、断言型、指令型、行為拘束型、表現型、宣言型という分類を提案した (Searle 1979)。これは、発話内目標、適合方向、誠実性条件という 3 つの基準に基づいた分類である。

さて、発話することが何かをすることであるのならば、言語行為は（他の行為と同様に）その道徳的性質を問うことができるのだろうか。仮にできるのだとすれば、その性質は、たとえば物を盗んだり人を殺害したりといった行為とはどのように異なるのだろうか。紙幅の都合でこの問題について網羅的に議論することはできないが、本論文の目的のために、ここでは発話内行為と（道徳的）規範の関係性について検討することにする。

オースティンの言語行為論のビジョンを体系化させたサール (1969; 1979) は、言語行為の道徳的性質を中心的に言及してはいないものの、事実と当為の関係について「自然主義的誤謬の誤謬」という、今日メタ倫理学でよく知られている議論を提出している。サールは以下の推論を取り上げ、記述的な陳述が実は評価的な陳述を帰結することを示唆する (Searle 1969, pp.315-316) ⁽⁵⁾。

1. 「スミスさん、あなたに 5 ドル払うことを私はこの言葉において約束しま

す」という言葉をジョーンズが発した

2. ジョーンズは、スミスに対して 5 ドル払うことを約束した
3. ジョーンズは、スミスに 5 ドル払う義務を自分に課した（引き受けた）
4. ジョーンズは、スミスに 5 ドル払う義務がある
5. ジョーンズは、スミスに 5 ドル払うべきである

上の推論では、「約束」の発話内行為を取り上げ、ジョーンズが発話したという事実からジョーンズがスミスに対してお金を払うべきだという当為を導くことの可能性を示唆している。サールはこれらの推論が論理的帰結（*entailment*）という必然的関係だとまでは主張していないが、それでも偶然的関係だけだともいえないと主張している。

なお、これは「約束」という言語行為について「約束は守らなければならない」という外在的規範が存在しているからこそ成立する、という話ではない。仮に「約束は破られるべきである」という規範が存在することを仮定してこの規範性を批判しようとしても、それは「約束」という発話内行為に関する外在的な批判であり、「約束」における内在的な言語制度——つまり義務を負うことそのもの——は保たれる。このように、発話内行為が遂行される時、そこには内在的な言語制度があり、その制度によって私たちは規範的な価値を背負うことになるのである。

サール自身はこの議論を倫理学の領域で積極的に検討しようとはしていないが、最近ではクーネオが道徳的实在論の立場を取る際に言語行為論の枠組みを利用し、発話における規範的側面を道徳的性質につなげようとしている（Cuneo 2014, p.77）。クーネオによれば、言語行為は規範的現象でありかつ道徳的現象でもある。彼の言語行為の規範理論においては、言語行為は発話者が発話者としての道徳的な権利や責任、義務を有しているという事実によって発話内行為とみなされる（Cuneo 2014, p.29）。つまりクーネオの立場では、道徳的価値は発話内行為の遂行における付随的な事実ではなく本質的な事実なのである。発話内行為と道徳的規範の関係性のあり方についてはさらなる議論が必要だが、少なくとも発話者が適切に発話内行為を遂行した事実を道徳的に価値中立的な事実だとみなすのは難しいと思われる。

技術的人工物の発話内行為の可能性

前節で確認したように、発話内行為を遂行する行為者はそれがいかなるタイプの発話内行為であれ、何らかの道徳的規範を伴うことがわかった。しかし、クーネオのような強い主張をするにせよ、そこまで踏み込まないにせよ、少なくとも発話内行為が適切に遂行されなければ道徳的価値は伴わないことになる。たとえば、人間の会話参加者が発話を言い間違えたり聞き間違えたりした場合や、動物の鳴き声が特定の発話内行為のように聞こえた場合においては上記の議論は関係がないことになるだろう。では技術的人工物がしかるべき計算（**computation**）によって「発話」した場合、それは発話内行為といえるのだろうか。この問題について、ここでは発話内行為における「慣習（**convention**）」および適切性条件に焦点を当てて検討したい。

オースティンが発話内行為の遂行において重視しているのが「慣習」の存在である。発話内行為にはそれぞれ当該行為を成立させるための慣習が存在し、それに沿った形で発話されなければ不発に終わる。たとえば「命名」であれば命名する権限をもっていることが要請されるし、「解雇」であれば解雇できる立場に立っていることが要請される。

では、発話する技術的人工物はこの慣習に関わることができるのだろうか。これに関しては可能だと思われる。まず、慣習というものは人間にとっての言語的・社会的慣習のことである。たとえば、教会で挙げられた結婚式において司祭が二人の婚約の成立を宣言することができるのは、結婚にまつわる社会的慣習が存在するからこそである。ここで、人間が、人間的慣習と関わりのない存在者の世界に人間的慣習をもちこんだとしても、発話内行為が遂行されたとはいえない。オースティンに従えば、たとえば、司祭が自室に飾ってある一对の人形に対して結婚を宣言することはできない。人形は人間の形をしているが、社会的存在とはいえず、人間の結婚に関する慣習を有していないからである。

しかしながら、人間以外の存在が人間同士の慣習に入っていくことは必ずしも不可能ではない。「約束」、「宣言」、「感謝」といった発話内行為に関する言語的・社会的慣習が現に存在しており、それに沿って技術的人工物が振る舞うことは可能だからだ。たとえば、銀行の **ATM** は取引終了後に「ご利用ありがとうございます」と「発話」するが、これは「感謝」に関する人間の慣習に沿った形で技術的人工物が行動する例だといえる。実際ここでは銀行側の（つまり人間の）感謝の気持ちを **ATM** に代弁させているわけだが、来客に取引してもらったことに対する感謝という点において、ここでの感謝はきちんと人間の言語的・社会的慣習に則っているといえるだろう。

以上のように、慣習の存在を理由に技術的人工物の発話内行為が原理的に不可能だと主張することはできない。とはいえ、慣習は発話内行為の遂行において重要な前提の一つではあるが、そのすべてではない。発話内行為の遂行におけるもう一つの重要な条件として、「適切性条件(felicity conditions)」がある。サール(1969)は発話内行為が適切に遂行されるための条件(適切性条件)を以下のように列挙している⁽⁶⁾。

- ・ 正常入出力条件：言語行為の遂行に必要な基本的条件
- ・ 命題内容条件：それぞれの言語行為の命題内容に関する条件
- ・ 事前条件：言語行為の本質ではないが、言語行為の成立に必要な条件
- ・ 誠実性条件：言語行為の誠実な遂行のために必要な条件
- ・ 本質条件：言語行為を成立させるために本質的な条件

たとえば「助言(advice)」において、上記の条件は次のようになる(S: 話し手、H: 聞き手、A: 助言)(Searle 1969, p.126)。

- ・ 命題内容条件：Hの未来の行為A
- ・ 事前条件1：SはAがHに対して利益を与えることを信じる理由をもつ
- ・ 事前条件2：HがAを行うのはSとHにとって自明ではない
- ・ 誠実性条件：SはAがHに対して利益を与えると信じている
- ・ 本質条件：AがHの最大の利益になるという趣旨を引き受けることとしてみなされる

では、技術的人工物はこれらの諸条件を満たすことができるだろうか。まず、命題内容条件はその名のとおり発話の中の命題の内容に関わる条件のため、適切なプログラムを組むことで満たすことはできるだろう。

本質条件についても基本的には満たすことができると思われる。本質条件を満たすためには、発話者が特定の発話をし、それが当該の条件を満たしているとみなされる必要がある。その意味で、技術的人工物は一見すると本質条件をクリアできないように思われる。しかし、この「みなされる」という表現がポイントで、本質条件を満たすかどうかは客観的な判断というよりはむしろ、聞き手による主観的・相対的なものである。そしてこの考えにしたがえば、技術的人工物は特定の本質条件を満たす可能性も否定できないと思われる。たとえば、携帯電話ショップでコミュニケーションロボット“Pepper”が来客を礼儀正しい「挨拶」で出迎えたとき、それは挨拶の本質条件を満た

しているだろうか。ここで問われている条件はあくまで「話し手が聞き手を礼儀正しく認知したこととしてみなされる」かどうかである。つまり、来客が“Pepper”の姿を見て、自分のことを礼儀正しく認知しているのだと感じたのならば、本質条件は満たされている可能性がある。よく知られているように、私たちにはコンピュータを社会的存在とみなしてしまう心理的傾向があるため (cf. Reeves and Nass 1996)、事前条件（この場合であれば「挨拶」が正しいタイミングでなされていること）が満たされている場合、本質条件も満たされていると考えることもできるのである。

しかし、それ以外の条件については、技術的人工物が文字通りに満たすことができるかと主張するのは難しい。というのも、正常入出力条件においては発話者の発話内行為に必要な言語能力が求められており、誠実性条件と（一部の）事前条件においては発話者の特定の心的状態が求められているからである。たとえば、「感謝 (thank)」においては、発話者が感謝の遂行に必要な言語能力を有していること（正常入出力条件）、感謝の対象が発話者に対して利益があることを発話者が「信じている」こと（事前条件）、発話者は実際に対象に感謝していること（誠実性条件）ことが求められるのである (Searle 1969, p.125)。

技術的人工物がこれらの条件を（文字通りに）満たすかどうかを考えることは、技術的人工物が（少なくとも特定の）言語能力および心的能力を有することができるかどうかという古典的な難問に向き合うことにつながる。ここではこれらの問いに取り組むことはしない。というのも、発話内行為の遂行者になるための条件として言語的・心的能力という関係性は、ロボット倫理学における道徳的行為者になるための条件としてしばしば心的能力が要請されることと同じ関係性だからである。この点を見逃して議論を進めることは、ロボット倫理学における近年の議論の蓄積を蔑ろにして既になされた論争と同じことを繰り返す「車輪の再発明」になってしまうだろう。したがって、ここではロボット倫理学における道徳的行為者性に関する議論と、本節においてこれまで行ってきた議論との対応関係の示唆に考察を留めることにする。

これまでのロボット倫理学においてはロボットの道徳性を論じるにあたって、ロボットが道徳的行為者 (moral agency) か否かが中心に議論されてきた。この論点について具体的にどのような立場があるかについてはガンケル (2012) がまとめているためここでは省略するが、このようなロボット倫理学の傾向はそもそもこれまでの倫理学が行為者性指向的であることが起因していると考えられる (Floridi 1999; Gunkel 2012, p.88)。この考えのもとで

は、ロボットの道徳的重要性を主張する際にはそれが道徳的行為者であると主張しなければならなくなる。

道徳的行為者性の要件としてこういった概念を含めるかは統一的な見解はないが、自由（意志）と特定の心的能力（自律性、意識、志向性など）が含まれる点においては概ね共通している。たとえば、自然に呼吸をすることと文字をタイピングすることの違いは、その背後の「意図（intention）」の有無の違いとして理解される。そして後者のような行為を遂行できる存在が、道徳的行為者だといえるのだ（cf. Himma 2009）。したがって、ロボットが道徳的行為者か否かという問いは、ロボットが（哲学的な意味で）自律的に行動を起こせるのか、自分自身の行動の道徳的な善悪を理解しているかといった心の哲学でおなじみともいえる議論に帰着することになる。

しかし、上記のような基準においてロボットが文字通りの道徳的行為者であるか否かを議論することには多くの困難が伴う。道徳的行為者性に求められる条件に含まれる概念は往々にして曖昧なものであり（Gunkel 2012）、また仮にそれらの概念が曖昧ではない形で定義できたとしてもその条件が満たしているかどうかは原理的に知ることができないといういわゆる「他者の心」の問題が立ちだかっている（Coeckelbergh 2009a）。直観的には現在のロボットは人間と同等の心的状態をもたないと思われるが、決定論的システムを本物の道徳的行為者と考えられるのかどうかという問題は、人間が本当に自由意志をもつのかどうかという問題と同様に答えが出ないものなのだ（Wallach and Allen 2008, p.84）。

一方で、この問題について、自由や自律性、志向性といったような概念を拡張し、ロボット（や他の技術的人工物）が何らかの意味において道徳的行為者だとみなせる、と主張する立場も多数存在する。フーコーを援用し自由概念を拡張しつつ技術的志向性の概念を持ち出すフェルベークの媒介理論（Verbeek 2011）や、適切な「抽象化のレベル（level of abstraction）」を設定することによって行為者であることのハードルを下げるフロリディとサンダース（Floridi and Sanders 2004）などはその代表例といえる⁽⁷⁾。

これらの「拡張」のアプローチの詳細な比較検討についてはガンケル（2012）が行っている。しかし、これらのアプローチのほとんどは、実際に意識や自由であることとそのように見えることとの関係性についての論争に行き着いてしまう。そしてこの論争を技術的人工物の発話の文脈でしようとするれば、「中国語の部屋」（Searle 1980）の論争の焼き直しになってしまうだろう。

一方で、ジョンソン（2006）のように道徳的行為者性の概念を拡張することなく技術的人工物の道徳的重要性を説明しようとする立場もある。彼女は

コンピュータについて、それが道徳的行為者にはなりえないが道徳的存在 (moral entities) ではあると主張する。岡本 (2012) が指摘するように、もし道徳的行為者性概念を使わずにその道徳的重要性が説明できるならば、オッカムの存在論的節約という意味でも望ましい。本論文では、少なくとも“Tay”のような技術的人工物の発話の道徳的影響力を説明する上では、以上で紹介したような道徳的行為者性概念に頼る必要がない、という立場を次節で展開する。

発話の媒介的性質とその道徳的重要性

これまでは言語行為の行為の分類の中でも発話内行為に着目し、発話内行為には道徳的規範が伴うことを確認した。また、発話内行為の適切性条件においてはしばしば発話者の言語能力や心的能力が問われるため、それを技術的人工物が遂行できるかを問うことは「技術的人工物は道徳的行為者たりえるか」というロボット倫理学やコンピュータ倫理学において長く続く論争に行き着いてしまうことを指摘した。

前節の最後で私は、技術的人工物が文字通りの道徳的行為者とはいえないからといって、道徳的行為者性概念を拡張する必要は必ずしもなく、道徳的行為者であることと道徳的重要性 (影響力) があることを独立的に考える立場を紹介した。ロボット一般、コンピュータ一般におけるこの立場の妥当性はともかく、少なくとも対話システムのような技術的人工物の発話においてはこのような立場を取ることが望ましいと思われる。最後に本節ではこの点について、発話媒介行為のほうに焦点を当てることで検討する。

サルやクレーネオの考察に従えば、発話内行為は道徳的規範を伴うものであった。しかし、だからといって、発話の道徳的性質は発話内行為のみによってもたらされるわけではない。発話内行為がどのような形で遂行されるにせよ、その発話によってもたらされる行為、すなわち発話媒介行為 (perlocutionary act) が聞き手の倫理的感受性に影響を与えることもある。

たとえば、ヘイトスピーチのように発話によって聞き手に屈辱を与えることについて考えよう。オースティンは発話媒介行為を説明する際に「屈辱を与える」という行為について言及している (Austin 1962, p.196)。ここで重要なのは、屈辱を与えるという行為には発話内行為のための定形表現が存在しないということだ。約束や命名といった発話内行為の場合、「私は(ここに) A する」という発話の形式を取ることができる。つまり、発話内行為には顕在的な行為遂行形式が存在する。しかし、「屈辱を与える」という行為の場合、「私は(ここに) あなたに屈辱を与える」という発話形式を取ることができ

ない。オースティンは「屈辱を与える」、「仰天させる」、「驚かせる」といった行為については目的よりもむしろ後続事件をもつ発話媒介行為として分析しているのだ⁽⁸⁾。したがって、発話内行為だけではなく、発話媒介行為とそれに伴う後続事件も同様に道徳的に価値中立なものではないのである。

一方、発話が聞き手に与える心理的効果だけではなく、その発話が聞き手以外の第三者の耳に入ることや、その発話がより広い言説空間に影響を与えるということも無視はできない。言葉の暴力を構成する感受とは、当事者のうちに生じうる過酷で鋭利な痛みばかりではなく、第三者や傍観者であっても感じ取らざるをえない怒りや居心地の悪さを含むより広く漠然としたものでもある（飯野 2016）。またバトラーが『触発する言葉』において指摘したように、発話にはそれが発話者に留まらずに「再演」されるという側面がある（Butler 1997）。いかなる発話においても発話者だけがその起源なのではなく、その発話者の前後に意味作用の繋がりが広がっており、発話の反復と再意味付けがその発話者を超えて行われているのだ（Salih 2002）。したがって、発話の道徳的重要性を考察することにおいては、発話内行為に伴う道徳的規範の問題だけでなく、発話〈によって〉なされる行為や影響にまで分析の射程を広げなければならないのである。

そして、以上のような発話媒介行為の道徳的側面に関する議論を技術的人工物に応用することは十分に可能であるように思われる。なぜなら、発話媒介行為は発話内行為と同じく「行為」という名前が付いているが、実際その性格はかなり異なるからだ。発話内行為の遂行可能性については（すでに見たように）慣習と適切性条件が重要だが、発話媒介行為においてはこれらの要素が必ずしも関わってくるとは限らない。

「慣習」については、オースティンも指摘しているように、発話媒介行為の遂行においては、発話内行為のそれに相当する慣習は存在しない。たとえば、「驚かす」という発話媒介行為が行われるためには「後ろにクマがいる！」と警告してもいいし、単に「危ない！」と叫んでもよい。つまり、驚かすという発話媒介行為の本質は聞き手が驚くという心的状態をもつことにあるため、そこに至る道は無数に存在するというわけだ。また、同じ発話内行為が異なる文脈で全く異なる発話媒介行為を引き起こすこともある。個別の発話内行為が慣習に縛られた形で遂行されるのに対して発話媒介行為は慣習と必然的には繋がっていないのである。

同様に、「適切性条件」についても発話内行為のような厳しい条件は存在しない。技術的人工物の発話内行為について議論する際には言語能力が問われる正常入出力条件や、心的能力が問われる（一部の）事前条件や誠実性条件

が問題となった。発話内行為は現代的な行為論や倫理学で議論される「行為」の一種だからである。一方、発話媒介行為においてはこれらの条件は必ずしも必要ではない。たとえば、ネコは人間に匹敵する言語能力はもたないが、その発声によって聞き手の私たちが癒やされるということは十分ありえる。同じように、「後ろにクマがいる！」という発話によって「驚かす」という発話媒介行為が遂行されたとき、話し手は驚かす意図があってもなくても構わないのである。発話媒介行為は「行為」という名前とは裏腹に、その内実は、話し手から聞き手に伝えられた心理的「効果」のことを指しているのだ。

以上で検討してきたように、発話媒介行為を遂行する「行為者」はロボット倫理学において通常想定されている道徳的行為者とは異なる内実をもつ。そして、発話の道徳的重要性は発話内行為にもあるが、その発話によって行われる行為（すなわち発話媒介行為）にも存在する。また、発話のもつ力はそれが「再演」されることで増幅するという側面もある。したがって、たとえば技術的人工物が発話内行為を遂行する行為者として不適格だとしても、それによってただちに技術的人工物の発話の価値中立性が帰結されるわけではなく、その人工物の発話がユーザや社会に与える媒介効果およびその発話の「再演」にも分析の焦点を当てなければならないのだ。

結論

本論文では、対話システムやコミュニケーションロボットのような技術的人工物の発話の道徳的重要性について、言語行為論の観点から分析した。従来のロボット倫理学においては技術的人工物の道徳的重要性と道徳的行為者が密接に繋がっていたため、技術的人工物が道徳的行為者か否かが重要な問いとして扱われてきた。一方で、ジョンソンのようにその重要性を行為者性から分離して議論しようとする立場もあり、少なくとも技術的人工物の発話においてはこの立場は妥当であるように思われる。技術的人工物の発話は、発話内行為、発話媒介行為、そしてそれが当事者以外の存在に与える影響という異なる観点から分析することが可能であり、これらのすべての要素がそれぞれ道徳的重要性をもつ。技術的人工物の行為者性を問う従来の議論はこれらのうち、技術的人工物の発話内行為の可能性を問うものであった。したがって、仮にこの議論が答えの出ないものだとしても、私たちは技術的人工物の発話について、それがもつ発話媒介的な効果およびそれが「再演」されるという観点からその道徳的重要性を説明することが可能なのである。

注

- (1) 本論文ではハードウェアをもたないプログラムについても「技術的人工物」の中に含めることにする。
- (2) この事例の詳細についてはネフとナギ (Neff and Nagy 2016) を参照。
- (3) ロボット倫理学の代表的な論文集としてはリンら (2011; 2017) がある。
- (4) 言語行為論の枠組みが対話システム開発においてどのように応用されているかについては中野ら (2016) および石崎と伝 (2001) を参照。
- (5) 本論文では文献のページ数を記載する際、邦訳が存在する場合にはその邦訳を文献欄に明記した上で邦訳版のページ数を記載することにする。
- (6) 適切性条件の各条件の要約は片桐 (1995) を参考にした。
- (7) ただし、これらの研究において、道徳的行為者の概念を拡張することによってなにを目的としているのかについては必ずしも一致していない。たとえばフロリディとサンダースの立場においては道徳的行為者性と責任概念が独立して語られている。技術的人工物の発話における道徳的行為者性概念や責任概念、その他道徳的価値の関係性については別のところで論じる。
- (8) ただし、発話内形式の不在からただちにそれが発話内行為ではないと帰結するわけではない。たとえば飯野 (2016) は侮辱のような特定の言葉の暴力が発話そのものと切り離せない関係にあることを指摘した上で、「侮辱する」は「ほのめかす」といった行為と同様に特殊なタイプの発話内行為ではないかと示唆している。

文献

- Austin, J. 1962. *How to do things with words*, Oxford University Press.
(= 坂本百大訳 (1978)『言語と行為』, 大修館書店) .
- Butler, J. 1997. *Excitable speech: a politics of the performative*, Psychology Press. (= 竹村和子訳 (2004)『触発する言葉』, 岩波書店) .
- Coeckelbergh, M. 2009a. “Virtual moral agency, virtual moral responsibility: on the moral significance of the appearance, perception, and performance of artificial agents,” *AI & Society*, 24.2, pp.181-189.
- Coeckelbergh, M. 2009b. “Personal robots, appearance, and human good: a methodological reflection on roboethics,” *International Journal of*

- Social Robotics*, 1.3, pp.217-221.
- Cuneo, T. 2014. *Speech and morality: on the metaethical implications of speaking*, Oxford University Press.
- Floridi, L. 1999. "Information ethics: on the philosophical foundation of computer ethics," *Ethics and Information Technology*, 1.1, pp.33-52.
- Floridi, L. and Sanders, J. 2004. "On the morality of artificial agents," *Minds and machines*, 14.3, pp.349-379.
- Gunkel, D. 2012. *The machine question: critical perspectives on AI, robots, and ethics*, MIT Press.
- Himma, K. 2009. "Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent?" *Ethics and Information Technology*, 11.1, pp.19-29.
- Johnson, D. 2006. "Computer systems: moral entities but not moral agents," *Ethics and Information Technology*, 8.4, pp.195-204.
- Lin, P., Abney, K. and Bekey, G. (eds.) 2011. *Robot ethics: the ethical and social implications of robotics*, MIT Press.
- Lin, P., Abney, K. and Jenkins, R. (eds.) 2017. *Robot ethics 2.0: from autonomous cars to artificial intelligence*, Oxford University Press.
- Neff, G. and Nagy, P. 2016. "Talking to bots: symbiotic agency and the case of Tay", *International Journal of Communication*, 10, pp.4915-4931.
- Reeves, B. and Nass, C. 1996. *The media equation: how people treat computers, television, and new media like real people and places*, Cambridge University Press. (= 細馬宏通訳 (2001) 『人はなぜコンピュータを人間として扱うか:「メディアの等式」の心理学』, 翔泳社).
- Salih, S. 2002. *Judith Butler*, Routledge. (= 竹村和子訳 (2005) 『ジュディス・バトラー』, 青土社).
- Searle, J. 1969. *Speech acts: An essay in the philosophy of language*, Cambridge University Press. (= 坂本百大, 土屋俊訳 (1995) 『言語行為: 言語哲学への試論』, 勁草書房).
- Searle, J. 1979. *Expression and meaning: studies in the theory of speech acts*, Cambridge University Press. (= 山田友幸訳 (2006) 『表現と意味』, 誠信書房).
- Searle, J. 1980. "Minds, brains, and programs," *Behavioral and Brain Sciences*, 3.3, pp.417-424.

- Verbeek, P. 2011. *Moralizing technology: understanding and designing the morality of things*, University of Chicago Press. (= 鈴木俊洋訳 (2012) 『技術の道德化：事物の道德性を理解し設計する』, 法政大学出版局).
- Vinyals, O. and Q. Le. 2015. “A Neural Conversational Model,” arXiv preprint, arXiv:1506.05869.
- Wallach, W. and Allen, C. 2008. *Moral machines: teaching robots right from wrong*, Oxford University Press. (= 岡本慎平, 久木田水生訳 (2019) 『ロボットに倫理を教える：モラル・マシーン』, 名古屋大学出版会).
- 飯野勝己. 2016. 「侮辱と傷つけることー「発語内の暴力」序説ー」, 『国際関係・比較文化研究』, 14.2, pp.1-22.
- 石崎雅人, 伝康晴. 2001. 『談話と対話』, 東京大学出版会.
- 岡本慎平. 2012. 「人工的道德的行為者への近年のアプローチの検討」, 『HABITUS』, 16, 西日本応用倫理学会, pp.53-73.
- 片桐恭弘. 1995. 「エージェントのメンタルモデル」, 『人工知能学会誌』, 10.5, pp.668-676.
- 中野幹生, 奥村学, 駒谷和範, 船越孝太郎, 中野有紀子. 2015. 『対話システム』, コロナ社.
- 東中竜一郎. 2016. 「対話システムと倫理」, 『人工知能学会誌』, 31.5, pp.626-627.

※本研究は、JSPS 科研費 JP18J23409（日本学術振興会特別研究員奨励費）の助成を受けたものである。