

オンライン・ヘイトスピーチ規制の法内在的制約（2）

－ 発語内的な規範設定の特質とその阻却可能性

The Regulation of Hate Speech and Inherent Limitations(2)

: On the Nature of Illocutionary Norm-Enactment and Defeasibility

永石 尚也*

Naoya Nagaishi

はじめに

オンライン上のヘイトスピーチへの規制の議論ではしばしば、リアル空間でなされた場合にヘイトスピーチだと認識される典型的発話が（SNSや動画共有サイト、あるいは掲示板フォーラム等、何らかの形態の）オンラインでなされた状況を想定し、その上でリアル空間における規制の類推や拡張の可否や、それに即した規制の諸条件が論じられてきた。しかしこの種の議論は、現実の排除や「貶め」等の各種の機能から定義されるはずのヘイトスピーチについて、発話内容のタイプから特定する前提を導入している点において、そもそも概念的に誤りを含んでいる¹。この結果として、「現実の排除や「貶め」とはかかわりのない表現をも含んだ規制をも正当化する点において、実践的にも誤導的のものになってきた。

そこで本稿では、以下の三つの課題に取り組む。第一に、有害と目されるオンライン上の発話を持つコミュニケーション・空間的な特質を

踏まえつつ、会話において許容される事実を設定する発語内行為とその効果に着目する分析の射程を明らかにする。特に、会話において許容される事実を設定するメカニズム（状況的規範の設定メカニズム）が、オンライン上では異なる機能を持つことに着目することで、具体的な対抗言論組成のバリエーションについて示唆を得る。

第二に、本稿ではオンライン空間上で有害と目される情報へのアクセスがなされる場合のコミュニケーションの多くは、（即時的な発話と応答の連鎖からなる会話状況とは異なり）通時的に蓄積されていく情報への積極的な参照と切り離すことができず、またヘイトスピーチとしての機能の実現についても発話時点以降になされる「物言い」に依存する傾向がリアル空間に比して高い点に焦点を当てる。この環境下においては、ポジティブカウンタースピーチおよび通時的カウンタースピーチとして議論される対

* 東京大学大学院情報学環

キーワード：ヘイトスピーチ、言語行為論、コンテンツモデレーション、対抗言論、スコアキーピング、フィクションの真理

抗措置が発話の危害を抑制する観点からは有効となり、その具体的な形態として発話の持つ暗黙の前提を可視化する「ブロッキング」戦略、当該前提をより平等主義的なものへと受け手側が曲げて解釈することで、元の発話の意味を修正する「辻合わせ bending」戦略、発話の効力を宙吊りにする「非決定」戦略が有用となりうることを示す。上記の戦略は、特定の発話の意味・意図の認定に基づく規範逸脱的発話の禁止や制裁付与、あるいは個別的な投稿の削除や発話者のプラットフォームフォーミングなどと異なり、元の発話の効力を明示的あるいは黙示的に限定・修正する点において、意味・意図の認定にまつわる困難を避けつつ、発話の効力を規範適合的なものとして保持させる特徴を有する。この点において従来の規制手法に比して優位性を示すことを、併せて示す。

第三に、上記の作業を通じて、有害と目され

るオンライン上の発話に対する諸規制のバリエーションのうち、現在の情報環境下において法的に許容しうる規制態様とその条件を定式化することを目指す。

悪しき差別につながりうる深刻な害は少ないに越したことはないものの、個別的な害を除去しさえすれば他の問題がどうでも良くなるわけではない。特定の社会状況においては自明に思える害を規制によって除去しようとする場面においてこそ、当該除去手段としての規制とその濫用のおそれに伴う社会的コストを正確に把握する必要がある。本稿を通じて、オンライン空間における認識上の限界を踏まえた有害な発話への実践的対応について、オンライン上のなりすましの問題への対応や偽情報規制の難しさとの比較を通じて、オンライン上のヘイトスピーチ規制への今後の展望を示す。²

1、オンラインヘイトスピーチの発話内行為としての側面

1.1 スコアキーピング（スコア記録）と s 規範設定

さて、前稿ではオンラインヘイトスピーチと目される発話が引き起こしうる害、すなわち発話と因果的に結びついた特有の害を特定するアプローチには種々の困難がつきまとう事態を明らかにした。この困難の原因は、個別の発話が因果的にもたらす害が、表現が置かれた文脈と表現の効果を遮断する主体やアーキテクチャの作用によって複雑化しがちであり、またその実現が不確実なものとなることに求められる。すなわち、発話がその後に続く事態の推移に依存しているために、因果的な偶発性を伴う点にそ

の問題を抱えている³。オンライン空間においては、発話が引き起こす (cause) 害は、その匿名性や高度の流通性、情報の持続性などの性質によって増幅されるばかりでなく限定もされる⁴。たとえば SNS 上のテキストは、発話が突如として自身を襲うというよりも、自身の内言として再構成される過程を介してその効果を生じさせる特徴を有する。このように、リアル空間においては自らの住まう環境にヘイトスピーチに属する発話群が物理的に回避困難な形で介入してくる一方で、オンライン空間で完結する

有害な発話群については、この介入が間接的であり、技術的にも選択・遮断が可能である場合が多いことが想定される。

こうした見解に対して、オンラインヘイトスピーチを含む有害な発話は、続く会話において許容される事実の（例えば平等性などの規範に反する形で）設定を直接になす（do）ことで、害を構成するとする見解がある。この見解は、発話をなすことによって（by）生み出される害というよりも、発話をなすことにおいて（of）構成される害に着目する。この結果、特定の状況下においてヘイトスピーチの発話それ自体が続く会話で許される発話や真となる発話を指定（制約）する規範を作り出し、当該規範の瀰漫によって、社会全体の差別的な環境を維持・助長すると捉える。

以下ではこの見解に立つマクガワンの議論⁵を批判的に検討した上で、同理論を（その多くが典型的な会話状況からは逸脱する）オンラインヘイトスピーチへと適用できるかについて詳述する。

マクガワンによる議論は、次のように要約できる。リアル空間における会話状況においては、ある発話が伝達する内容に加え、発話それ自体を真あるいは妥当なものとするように発話の背景をなす文脈を調整（調律）する「受け入れ（accommodation）」の力が、当該発話の瞬間に発動する。この受け入れは、会話が可能な限り滞りなく調和的に継続されること自体へ貢献するように誘因する同調作用による。この結果、調和の達成を会話参加者自らが評価する基礎となるスコア、すなわち、「会話スコア」に動的に変更が生じる。つまり、続く会話の進行

の中で許容される個別の発話が調整（調律）され、その調整（調律）により変更された新たな会話スコアに照らして、続く発話の真偽や許容性の判断基準、その適切性条件が規範的に設定（enactment）される。これらの規範は状況的規範としての「s 規範」と呼ばれ、より広範な社会的な構造を反映した基盤的規範としての「g 規範」に支えられることで設定される一方で、また逆に g 規範の維持にも寄与する。こうして設定された s 規範について、事後に異議を唱え、否定を試みるということが可能であるとしても、それは発話による発話内的な効力（例えば抑圧等）を事後的に取消すこと（undo）であって、一旦生じた効果を遡って無にすることではない。この効力のうち、s 規範が現実には特定の属性を持つ人々を排除したり、人々の平等な地位を貶めたりする機能を果たす場合、社会的状況と発話そのものの効果（force）⁶による害、言い換えれば、発話それ自体によって設定された規範による害や、その規範によって続く発話（典型的には異議や否定等）が抑圧される害が構成されることとなる。

以上が、典型的な会話状況における s 規範設定についてのマクガワンの議論の要約である。さらにマクガワンは、会話状況以外の社会的な規範に統制された活動においても、一度生じたスコア変更による s 規範設定は一般的に生じるものであるとも主張する⁷。つまり、非会話状況においても会話規範と同様に、抑圧をもたらすなど望ましくない規範設定⁸は遍在しているとマクガワンは考える⁹。特殊な地位や特殊な環境は我々の社会的相互作用を統制する g 規範の一部であり、特定状況下の s 規範設定・スコ

ア変動に関わる蓋然性は高いだろうものの、必然的關係にはない。体系的に差別的であったりする g 規範の一部を通じて、s 規範は場面ごと

に設定され、その集積が悪しき g 規範としての機能を維持させる循環をなす関係に立つ¹⁰。

1.2 ヘイトスピーチの害の構成と、続く発話による効力の限定

このマクガワンのアイデアを引き継ぎ、近年、ヘイトスピーチの害の構成メカニズムの分析やそれに基づく対抗的アプローチ態様・規制態様の分析において、この「受け入れ」がもたらす危害が論じられている¹¹。とりわけ、個別的 s 規範は、(広く了解された基盤的 g 規範と結びつくことで) 覆したり変えたりするよりも設定する方が容易だという、非対称な「粘性性」を持つとの指摘は広く共有されている¹²。この粘性性ゆえに、悪しき発話に対する異議や否定自体も、争点化された集団の属性を顕著なものとして際立たせてしまう点において問題があることも認識されている。マクガワン自身も明示的に、上記のアイデアを対面状況におけるヘイトスピーチに適用しており、さらに近時においてもこの立場は維持されている¹³。しかしこの立場には、理論上の難点及び具体的事例への適用上の難点が含まれている。

第一の難点として、マクガワンが自身の理論の基礎とした D. ルイスのアイデアと、マクガワン自身が提示した s 規範のアイデアとの理論的な衝突の問題がある。元々のルイスの論文(1979)におけるスコアキーピング(スコア記録)というアイデアは、言語を通じた様々な実践における「正しさ」の構築とその変容過程を取り扱うものであり、(マクガワンが捉えているような意味論的モデルというよりも) メタ意味論的モデルであると考えられる¹⁴。すなわち、

ゲームにおいて「正しい」とされるプレイとそうでないものを作り出したり、ある会話において「正しい」とされる前提や事実を作り出したりするスコアの運動が存在し、その運動の仕方(kinematics)を参加者や観察者が追跡しながらゲームや会話が進行するとともに、そこでのスコア関数に変容しつつゲームや会話における「正しさ」を統制するとするアイデアであった。

例えば、ある上下関係に基づく会話上の命令が(命令に続く行為として)許容される「正しい」行為の範囲を設定する。あるいは、公式の野球のゲームにおけるスコアボード上の記録が(記録に続く行為としての)打者や走者の「正しい」動きやその違反、あるいはゲーム進行としての攻守交代を統制する。これらと類比的に、会話における「正しい」とされる前提や指示対象、場所や時点の参照点、曖昧な発話における「正しさ」の境界やその精度、可能性を含む様相文の真偽、さらには結婚や命名などその遂行に際して社会实践に裏付けられた適切性条件を要する発話においては、「正しい」文は当該会話自体の進行と会話を含む社会实践に依存して変動することとなる。これらのスコア記録が続く発話における「正しさ」を設定するとともに、そのゲーム上の「正しさ」がスコア運動に伴い変動することとなるのである¹⁵。

さらに、記録されるスコア自体さえも不確定なまま宙吊りにされるゲームや会話もある。例

えば、緊急事態・犯罪の発生を想定した計画立案や組織継続に向けて一定の不確実性を伴う対策想定等に見られるように、当該会話自体が参加者自身による遂行を伴わない、複数の反事実的条件法 (counterfactuals) 使用とその留保を含みながら進行する会話なども想定しうるだろう。そこでは、会話の参加者自身が必ずしもその発話の「正しさ」に現実においてはコミットせず、あくまでも (非参加者としての) フィクション的な仮定としてのみ通用する「正しさ」にコミットすることとなる。すなわち、「正しさ」が通用するパターンを複数同時に保持した上で、会話進行上の諸仮定のうち強弱を持つもののみ通用させるスコア運動もまた存在することがわかる。例えば、現実の企業における予測に基づく行動計画を立てるためには、シミュレーションとして外的事情変更を組み入れた諸シナリオを同時に観察しつつ、それに応じた複数のモデルを介して行動選択を適時に調整する必要があるだろう。

これらに見られるように、ルイスの元々のアイデアは、「正しさ」が参加者を統制する側面を、かなりの程度流動的なものとして捉えており、会話の参加者として「正しさ」を付与し、互いを拘束する会話に限られない。それは公式・非公式の種別においても異なるし、会話の外側にある社会実践の種類や強度においても異なるだろう。会話自体が公式・非公式なものに分離すればその「正しさ」もまた変動する上、会話における社会実践や「正しさ」の種類によってもまた、その拘束の程度は異なる。例えば、事実を証跡する言語ゲームと当該会話上の整合性を構築する言語ゲームとではその運動の仕方

は異なるだろうし、潜在的な曖昧さをより明確にしていく言語ゲームではそれを「正しくする」力が生じる傾向を持つ一方で、計画立案などの (会話中に現れる諸スコアとそこから導かれる「正しさ」が、複数の可能なシナリオごとに、その切り替えの都度全体論的にしか定まらない) 言語ゲームにおいては「正しさ」のメタ的調整がより重要となるだろう¹⁶。

これに対して、マクガワンが想定している会話は、より単純化された会話に焦点を当てている。すなわち、一定の固定性を持つ会話の参加者として、その会話における「正しさ」の維持や遵守にコミットしている参加者により構成された会話に焦点を当て、そこからの一般化として会話状況以外の社会的な規範統制的活動に理論を拡張しようとするものである。しかし、これはルイスが想定したスコア記録におけるごく一部の運動の仕方を、過度に一般化したものである。スコア記録の運動が通用する場面は、マクガワンが想定したような、非争点的な前提を付与し、あるいは「より正しい」発話へと促す表現ばかりではない。実際、オンライン環境の中でも SNS プラットフォーム上の表現についてこれを見れば、特定の会話が始まってさえおらず、いつなされたか、誰に対してなされたか、誰ともになされているかのいずれもが不確定であるケースがほとんどであろう。マクガワンが想定している例はいずれも、何らかの (多くは現実の) 発話者の権威を前提に、会話参加者が共同の貢献を自ら選択的に実行する前提に立つものであるが、ルイスがあげている例のように、会話全体の中で見ればそうしたコミットメントが要求される会話は限定的であり、オンラ

イン空間においてはなおさらそうであるだろう。これらのゲームにおいては、発話時における前提導入や規範設定はなされず、発話の効果は宙吊りにされる傾向にある。

以上のように、マクガワンが提示した「s 規範設定」モデルは、ルイスの提示した例のごく一部をカバーするものでしかない¹⁷。さらに重要な点として、マクガワンの想定とは異なり、スコア変動が即時的になされるとは限らない点が挙げられる。実際、ルイスは、スコア変動は宙吊りや事後の阻却 (defeat) が可能であり、言語ゲームの種別ごとに異なるスコアの動きを記述する必要がある旨を明確に指摘していた¹⁸。この点につき、「受け入れ」を、文脈を措定することで続く命題の真偽やその妥当性を決定する意味論的な水準においてではなく、その真偽や妥当性の決定基盤の変動メカニズムを明らかにするメタ言語的な水準¹⁹から解釈するものが見受けられる。すなわち、「受け入れ」の作用により議論が継続し、スコアが徐々に固定されていくことで命題の真偽や妥当性 (意味) を作り上げる時間的プロセスに着目する見解である。時間的外在主義と呼ばれるこの見解によれば、スコア変動はむしろ、時間的に後の事象経過を取り込み、遡ってもとの発話がなしていたこと (do) がなんであったのかを固定させていくとする²⁰。

さて、第二の難点として、以上の理論的な問題が、具体的な事例への適用においても s 規範の判断とその適用の判断を困難にする点が挙げ

られる。スコア変動によって作りだされる s 規範は、会話が非典型になる程度に応じて、(不発や誤発とマクガワンならばいうだろう) 別の s 規範を惹起する。s 規範が話者らの心理的要素を超えた外的要素 (ルールなど) も取り込みうる以上、そこで s 規範が定型的に通用する (行為ルールをなす) ためには、その定型を形作るルール (構成ルール) が必要であるが、法定とは異なり、この s 規範設定は後者のルールの確定要素を欠いている。

仮に s 規範の害が実現するとしても、実現された s 規範は複数でありえ、その害を取り除くものが規制であることもまた帰結しない。つまり、話し手にもまして、聞き手は分散した形で成り立ちえ、設定される s 規範についてもまた多分岐化した別の s 規範の作用を生み出しうるのである²¹。s 規範の害が一義的ではなく生じる以上、会話の中で微小なコストを強いる日常的な沈黙化や抑圧一般は、質問や訂正によって介入可能な場合がほとんどである²²。

また実践的にみても、s 規範設定の多分岐性とスコア記録の進化的特性からすれば、s 規範設定そのものに過度な沈黙化や抑圧を読み込む見解は、むしろ行為原則の過重による行為選択上の「麻痺」を引き起こす²³。このように日常的な s 規範の害の多くは法的介入と結びつくというよりは、むしろ発話のより適切なスタイルの提案 (あるいはケア・共生の理念) と強く結びつく。

1.3. 集積・流通する言説の害と g 規範

以上の検討からは、s 規範が構成する害のメ

カニズムが、より広範な g 規範の内容や参与者

の属性に依存していることがわかる。そしてオンライン空間においては、g 規範の維持が、特定の情報流通環境においてアクセス可能な情報の正確性やその偏り、あるいはそうした言説状況への介入（修正や追加等）や対抗手段選択と強い関連を持つことがわかる。実際、アル空間のそれにおいて、「慣習的な言葉と文章の意味の核心」が、伝達の瞬間、解釈の余地なく伝わる一方で、オンライン上では、まさしくそのヘイトスピーチが作り出す敵対的な空気と距離を持っているのであり、「批判者として断固とした姿勢で立ち向か」わせる慣習的な力が機能しているとは言い難い²⁴。

このように、オンライン空間における個々の発話については、この環境がもたらす複合性の上で、表現の効果を測る必要がある。ここで、個別的発話が流通するオンライン空間自体の特性として近時論じられている、「情報の他律的摂取」やそこからの偏向の助長は、s 規範がオンライン空間で働くメカニズムの一つを示している²⁵。

個々の発話が集積し、流通する環境の改善（媒体ごとのインセンティブ設計／訂正手続き設計）と関わる上記の指摘は、オンライン空間とリアル空間とにおける差異としての主体の重量性と、対応の多層的構造とに着目する。これによれば個々の表現の持つ効力は、当該表現が流通する技術的環境（広義のアーキテクチャ）及びそれを担う主体（プロバイダ等）に依存することとなる²⁶。ここから発話類型としての違法・有害情報への対応にあたっては、削除のみならず、警告や（一時的）閲覧制限、情報開示、フィルタリング、デプラットフォーム等、様々

実現形態をもち、その選択にあたっては、メディア・技術の変化に即したサービスの実現はもとよりとして、利用者の表現の自由、プライバシー、通信の秘密等の法的権利・利益を支える役割をも担うこととなる。この結果、対応の実現手法としても事前のアーキテクチャ・事後の判断、そのプロセスとしての審査のありよう²⁷を含む直接的対応に加え、（事業者ではない）個々の訴訟における解決に向けられた間接的対応まで、幅を持つこととなる。

この特性から期待される法の役割もまた、表現に関与する事業者に直接に行為を義務付け、規律するとことには求められない²⁸。萎縮効果及び同調圧力と強く結びつく刑事的規律と市場におけるソフトロー形成・秩序形成を促す市場法的規律との違いを踏まえた対応が求められるのであり、事業者の取り組みにおけるアカウントビリティを確保し、適正化するべく（政府言論としての反差別の保障はもとより、議論フォーラム提供のための共同規制枠組みの組成等を含む）「先制的カウンタースピーチ」を通じた規範回復のための諸措置の配置が求められる。すなわち、ここではs 規範の阻却可能性とその期待、g 規範の平等化志向の発展とその期待とが、個別的状況を超えたアクターの組成と配置についての集合的決定と強く関連しているものといえよう。

マクガワンが念頭に置いていたように、むしろ会話がs 規範を受け入れて進み、訂正が不能になる状態というのは、s 規範違反を受け入れている蓋然性の高い状態が継続し、訂正がなされないままに関係者が退出する場面を想定しているように思われる。しかし、オンライン空間

の場合には、スコア記録を含めたやりとりをその訂正過程をも含めて記録化することは、(リアル空間における個別的会話とは異なり)アクターの幅に照らしても、また技術的にも可能である。ここから、発話の悪しき効力の限定のた

めに採用しうる手段として、そのスコアボードおよびスコア記録の運動の仕方においてなすことが、直接的な規制作用との比較の俎上に載ることとなる。

2、カウンタースピーチの発話内行為としての側面

2.1 なされた発話内行為の効力の取消しと、取消し自体の害の構成

以上の通り、悪しき s 規範設定が働く場面は限定的に特定される。以下では、仮に悪しき s 規範設定がなされた場合において、いかなる阻却メカニズムが働くか、より具体的には事後の取消し (undoing) の諸例について検討する。これらの取消し作用は、発話の事後になされる以上は s 規範設定の因果的な発話媒介行為の効力を遡って無かったことにするのではなく、典型的にはその基礎にあった発話内行為の効力をその訂正過程の中で (スコアとして記録しつつ) 取り消すこととなる。つまり、発話が意図的あるいは意図せずにもたらすことが想定されていた発話内行為の効力について、聞き手側の応答によってその成否を変動させうる対抗的な発話内行為として捉えられる。ルイスの枠組みに従えば、事後に発話のスコアを変動させ、元々の発話が足を置いていた基礎を失わせることで、新たなスコアとして記録するわけである。

さて、最も想定しやすい取消しの例は、「異議」の形態をとる。カウンタースピーチと呼ばれてきた典型も、有害な発話の内容に対して反論し、その妥当性を疑わしいものへと変じることを企図した発話である。しかし既に見た通り、カウンタースピーチはヘイトスピーチの粘

着性を十分に考慮に入れておらず、またカウンタースピーチによってほぼ争いのない争点をあたかも「現に争われている」ものとして顕在化させてしまう点がしばしば問題視されてきた。各種のヘイトスピーチについても、特定集団をその属性に基づき貶める表現が典型的にはなされる以上、その属性についての認識違いや評価上の誤りを指摘することは、却ってそうした論点が争われている最中にあるかのように提示し、記録してしまうというわけである。さらにカウンタースピーチが功を奏するためには、元の発話が真偽を争いうる陳述の形式をとっており、誠実にその真偽を巡って論争を行い、論争の結果として自身の信念と陳述を撤回する態度を共有しているという、熟議の空間が開かれていることを前提としている。しかし、ヘイトスピーチを含む有害な発話の多くについてはこの双方が欠けることが多く、そのためにカウンタースピーチの機能が限定的である点もまた指摘されてきたところである。

この点に対応するため、Langton による「ブロッキング」戦略のように、元の発話が暗黙のうちに伝えていること (implicit content) を明示化し、発話者に帰属させることで元の発話内

行為の効力を失わせる戦略がある²⁹。言い換えれば、争点を巡って争うのではなく、争点形成自体の悪しき意図や悪しき発話の意味を明るみに出し、その弁明を要求する戦略である。確かにこの戦略は、元々の有害な発話が真偽を問いつらい発話であっても、その前提を真偽を問いうる命題として提示することで、発話者を議論の場に呼び出す点において有用性を持つ。これに発して熟議の場が開き、新たな議論が開かれることもあるだろう。しかし、発話者の暗黙の内容を明らかにすることは、発話者の「顔」を潰すことに繋がりやすく、それゆえに発話者によるはぐらかしや意見への固執を生み出す点において、対応として十分とは言えないことも指摘されてきた。

こうした問題に対応する取消しの形態としてCaponettoは、既になされた発話に対して受け手側が曲げて解釈することで、元の発話が設定したs規範を別の形に修正する「辻褃合わせbending」戦略を提示している³⁰。言い換えれば、この戦略は暗黙のうちに害をなす発話に対して、より平等規範に沿ったものにする前提に叶うように曲げて解釈し、発話者に対して再提示することで、元の発話者に、発話の前提をなしていた偏見が聴衆に受け容れられないかもしれないという感覚を与える対抗的発話の形態である。具体的には、元の発話者が偏見に基づいて発話したのではなく、そうした発話への怒りの表明であるという前提や、そうした発話がなされている事実についての論評であるという前提を新たに導入することで、元の発話者に更なる（多くは平等規範に沿うだろう）発話を促す点に特徴がある。この促しに対し、元の発話者は、

元の発話の意図やその発話内効力の認識を開示するように、s規範が持つ「受け入れ」を介して迫られる。通常、この自身の発話の潜在的差別性を改めて明示する発話をなすこと（rejoin）は避ける圧力が働くことから、コメントの潜在的差別性を阻却する介入としての効力を持つこととなる。Caponettoはこれをあたかも発話者の発話が実際よりも偏見に満ちたものでないかのように振る舞う「振り」の一形態とし、それにより、より偏見に満ちたものでない改善された内容をスコアに入れ込み、s規範に導入する寄与を果たすものとして提示している。

この「辻褃合わせ」戦略は（リアル空間においてもオンライン空間においても）相当に有効となる場面が見込まれる。ただし、この言い繕い自体が発話者の主体性を捻じ曲げる歪曲（distortion）の害を構成しており、またよいように言い繕う側が先んじて元の発話内行為の持つ悪性を一方的に認定してしまう点において新たな害を構成する点も指摘されている。すなわち、受話者が与えた解釈を元の発話者に暫定的にせよ帰属させ、元の発話者が批准/正当化の矢面に立つことを求める点においては、状況的にかなり偏見や差別的な意図が明らかなケースを超えて「辻褃合わせ」が行使されることは、除去しようとした害以上の別の害を生み出しうる。とりわけ、オンライン空間においては、発話がなされた場合においてそれが誰に対してみられる状態に置かれているかは必ずしも明らかではなく、対話相手ではなく対話を傍観しているものに向けた意図やアピールの形態をとってなされる場合もしばしばある。この場合には、「辻褃合わせ」戦略自体が単なる誤読に陥るこ

ともあれば、攻撃の暗黙的指示を因果的に惹起することも問題とされるだろう。

こうした点を踏まえて、オンライン空間については「非決定のままの途絶」という戦略もまた有用なものとして提示されている³¹。これは、単なる沈黙によって消極的受容を表現してしまうことや、逆に明確な拒否・軽蔑の意思などを伝達する雄弁な (eloquent) 沈黙によって s 規範に対抗することではなく、元の発話がその発話の力を持たなかった環境を作出することをその内容とする。実装例は多岐にわたるものと想定されるが、少なくとも対抗者にとっては、公然と立ち向かう負担を回避するとともに、元の発話への一定の評価を会話進行における「重み付け」として表示・記録する戦略として現れ、

2.2 先制的カウンタースピーチと g 規範設定

これに加えて、(s 規範がその基礎を置く) g 規範を改善するカウンタースピーチの組成もまた戦略として提案されている。例えば Lepoutre は、設定された s 規範に対する各種の「取り消し」をなすネガティブカウンタースピーチに対して、より平等規範に沿ったポジティブカウンタースピーチが「粘着性」に強いことを強調している。その理由として、「粘着性」は g 規範こそが助長する特性であり、発話が呼び起こす g 規範のうちの一つである平等規範に沿った形で議論環境が整備されていることが、s 規範の通用範囲を限定することにつながる点をあげている。

このような Lepoutre の戦略は、カウンタースピーチは元々通時的 (diachronic) なものであり、個別的 s 規範設定の前後で利用される、

元の発話者にとっては、一定の応答がありうるとしても、続く議論からの撤退を許容し、発話者を自身の意図の開陳や「辻褃合わせ」による新たな前提への取込みの圧力から開放する戦略として現れる³²。他と同様、この戦略も現に言われたこと (saying) を除去するのではなく、現に言われたことの意味内容を確定することなく宙吊りにする形で、その悪しき発話の効力を減退・中立化させるものである³³。

以上のような発話内行為の s 規範設定を、介入とその記録によって個別的に限定づける戦略のいくつかを組み合わせ、またその技術的実装を幅を持った環境の中で達成することが、規制主体に対しては求められうるだろう³⁴。

時間的連続体としての言説集合の一部であるとの前提を持つ。曰く、個別的な s 規範設定やそれに伴うスコア変動自体を問題にするのとは異なり、通時的に見た場合のスコア変動範囲を限定したり、事後的にスコアを立て直したり、蓄積的なスコア割り当て自体を議論に乗せたりすることが、カウンタースピーチが機能する g 規範設定と過剰な s 規範制限のためには重要となる、とする。これは、「辻褃合わせ」戦略でもみたように、対抗者の意図にかかわらず s 規範設定それ自体が議論封殺・議論拒絶を含む「有害なカウンタースピーチ harmful counterspeech」を構成しうる以上、避けて通れない問題となる。

元々、カウンタースピーチの効果は、悪しき発話の持つ危害を除去すること自体ではなく、

その際の平等を現実において達成することになった。このことに、Lepoutre の戦略は向けられている。すなわち、カウンタースピーチは、公的な発話の空間において、平等な発話者としての地位と尊重を割り当てているかを問題にするのであり、s 規範やその取消しにおいて参照される g 規範が歪んでいる場合や、その歪みを

恣意的に運用する g 規範が通用する環境においては十分に機能しないのである。対抗者に対する「(悪しき) ヘイト認定」との水掛け論やそれに基づくキャンセル文化の問題を回避する戦略としても、先制的カウンタースピーチの有用性を認めうる。

2.3 オンライン上の s 規範設定 /g 規範設定とその限界 私人・DPF・国家の役割

以上で見てきたように、個別的な s 規範設定は日常のどこにおいてもなされうる。その悪さは、続く発話を制限することにより、発話者としての平等な地位を奪い、あるいは発話の信用性や価値を低め、さらには発話者への議論封殺・議論拒絶をももたらしうる点にあることも見てきた。しかし、こうした悪しき効力は、発話自体が一般的にもちうる効力でもあり、発話内容が一見してもつ内容によって、その効力の発動の有無が変化するわけではない。上記の Caponetto や Lepoutre の問題提起は、s 規範あるいはそれを支える g 規範における効力の限定や、より平等規範に沿った効力の増進にこそあった。

さらに、特定の発話がいかなる s 規範の内容を持っているかを認定するにあたっては、事前であれ事後であれ、固有の困難があることも確認してきた。発話の空間的・コミュニケーション依存性に鑑みれば、その認定自体が持ちうる「悪さ」を加味して、続く発話のパフォーマンスを評価しうる状態を確保し、広く合意された手続きの上でその評価・改訂を実践していく他はない。しかし、その事後的評価は発話をめぐる弁明や「言い繕い」、あるいは改めて自身の

立場を明示する発話といった事後の取組み過程においてなされるのであり、対抗者が専断的に発話の悪しき規範設定を認定することもまた、新たな発話者の権威を設定することとなろう。つまり、しばしば主張される「より多くの発話 more speech」としてのカウンタースピーチの重要性は、単に「批判者として断固とした姿勢で立ち向かわせる」しばしば過重なものともなりうる負担としてではなく、上述してきた s 規範設定の普遍性およびその悪さの認定をめぐる事後的な評価の困難から要請される、通時的な s 規範の生成・解釈・確定プロセスとして理解することができるのである³⁵。

さて、以上を踏まえたオンライン空間上における s 規範設定および g 規範設定についての介入について、その手法面と主体面から若干の展望を示しうる。すでに見たように、s 規範設定の観点から見たリアル空間とオンライン空間との重要な違いは、s 規範設定の変化を記録するスコアボード自体が物質的形態をとりうる一方で、プラットフォームの構築・管理に依存する点にある。この点を捉え、例えばデジタル・プラットフォームにおける対応として発話流通環境を通じたスコア変動を記録・可視化するスコ

アボード構築・管理作業への従事を要求することは、より多くの発話を通じた個別的 s 規範の修正を機能不全から救い出す対応として理解できるだろう。

とりわけオンライン空間上においては、発話内行為 (illocutionary) としての s 規範設定をなす (do) 前に、その発話 (locutionary) が発される時点 (say) において、捕捉やそれを通じた介入が技術的に可能である点が特徴的である。すなわち、リアル空間における発話が発話

時点の瞬間まで介入が原理的にできないとされてきたのに対し、オンライン空間における発話は、発話がなされる過程への介入を技術的に可能とする。この原理的な違いは (可能ゆえに) 介入を要請し、また介入を正当化するのではなく、むしろ介入態様の精緻化を要請し、それによりむしろ続く発話内行為としての s 規範の内容を確定させるオプションを媒介者たる DPF に対して与えることとなるだろう³⁶。

3、進行中の規範の中断・変更と、On/Off Record を担う新たな規制者

本稿で見てきた「辻褃合わせ bending」戦略や「非決定」戦略は、発話の効力を規範適合的なものとして修正しつつ維持する可塑性を有していた。他方、この規範適合性そのものもまた不断に批判に晒されることで、その規範の権威を解体し、再構築する余地に開かれていなければならない。規範適合性の要求自体がもたらす抑圧現象について最後に若干付言することで、結びに代えよう。

さて、ジュディス・バトラーはある箇所では、一見するとその時点においては (発話上の権威など) 適切性条件を欠く主張 (assertion) を敢えて行使する積極的な意義について述べている³⁷。曰く、現に発話をすることによってその失敗を晒し、失敗の是正を求め、将来の道筋を規定 (prescribe) する。そうすることで、自己承認的な (self-authorizing) パフォーマティブがなしうとする。

この発言は、ジェンダーの割り当てをめぐる問題を巡りなされたものであるが、とりわけ

「ジェンダーの観念は子どもがこの世に誕生する前に形作られ」、我々もまた常に「進行中のジェンダー規範の場の中に存在する」以上、「その割り当てに対する転覆や断絶が起こるとしても、それは権力の場の中で、歴史的に形成された制約の場面の中での闘いを通して起こる」ことにバトラーは注意を促している。これにより、あえて適切性条件を欠く主張をなすことがもつ発話内効力と、その規範上の転覆が持つ政治的意義を明らかにしようとしたものである。先行する、進行中の (確定困難な) 歴史性や権力、システムの複数性を踏まえ、発話上の権威を逸脱する形で初めて議論の中に入ることのできる一群の観念が存在する。これは被抑圧の歴史性や権力の不均衡等をその要素とするヘイトスピーチについても、当てはまるだろう。本稿で規範設定について論じてきた箇所の語彙に引きつけられれば、ここでの問題提起は、発話の適切性という現在の規範適合性だけに議論を基礎付けることで、特定の観念を議論可能な状態から

遠ざけてしまう営為の粘着性に向けられているものと解しうる。

以上の紹介は、どの主体がその害を認定する権限（あるいはマクガワンの言い方を借りれば発話上の権威）をもつとされるかが固定される状況への警句として読むことができる。歴史的背景や言説上の蓄積を背景として、特定の語や文の利用が「奪取」されることで、対抗的発話が「規範適合的に」封殺され、無化され、忘却され、また書き換えられる種々の「抹消」は、歴史を通じて稀ではなかった。例えば「ユダヤ人の邪悪な憎悪が世界中を戦争、困窮、苦難に陥れた」と言う文は、ナチス側の人々に向けられた「プロパガンダ」であるが、同じ文をユダヤ人が見つけたら「攻撃」になる」とはラングトンのよく知られたフレーズであるが、もちろんこの受領者相関的な分析は、受領者の側に属する（と称する）者が「見つけた」瞬間において一方的に文の解釈を決定する権限を得る事態を示しているのではない（本稿の検討に従えば、それもまた「抹消」の一形態となる）³⁸。むしろ重要な問題は、真偽の問いつらい「ユダヤ人の邪悪な憎悪が世界中を戦争、困窮、苦難に陥れた」との曖昧な文が、その発話上の効力を受領者相関的に拡散させつつ、同時に機会主義的に否認可能な仕方ですべて使ってしまう発話が果たす規範設定の捉え難さと、この捉え難さを含めての対抗言論の新たな形態の組成と促進にこそある。

重要なのは、上記の文が「攻撃」としての解釈がなしうるものとして抵抗が可能な文であることを正当化するものであり、同時にそれ以上のものではないことを、一見したところの敵対

者と共有することである³⁹。抹消に抵抗する条件として抹消の記録があり、記録の可能性こそが敵対者との共有の基盤をなす。しかし実際にはこの解釈が「認定」へとスライドする諸局面を観察すれば、この「攻撃」認定がどの対象にかかり、どの対象にかからないのかの帰趨がかなりの程度、当該発話が置かれた具体的な文脈と時点に依存することがわかるだろう。論争的な語や文の意味内容について争われつつ、自らの言葉の用法を通用させようとする（メタ言語的）交渉が日々なされているSNS上での議論状況は、その議論環境の構築に寄与する介入形態の模索を迫るものである⁴⁰。

思えば、こうした濫用的言語使用による抑圧について、先ほど言及したジュディス・バトラーはインタビュー中で、自身のイスラエル国家批判が「ハマス支持的」なものとして誤読される解釈環境について語ると共に、そうした発話が「反ユダヤ」的なものとして一方的に規制の対象となる欧米諸国の一部における法執行環境自体に先取的に憂慮を示していた。「レトリックを駆使し、メディアを通じて、パレスチナをテロリズムと同一視し、全てのパレスチナ人をテロリスト・野蛮さ・獣性と同一視する。その際、同一視する人たちの想像の中には（パレスチナの）市民は出てこない。まさにその市民に向けて凶行を現に働いているというのに」⁴¹。

あるユダヤ人の一人によりなされたこの発話自体が、本章執筆時現在の一部の国において「ハマス支持的」なものとして誤読・拡散され、あるいは規制の対象となるのだという脅しがかけられていることを、文字通りオンライン空間におけるスコアボードに「記録」する必要がある。

しかし、この「記録」及びそれへのアクセス可能性や「転覆」を含む介入可能性が、いずれも新たな規制者としてのプラットフォームの振る舞いに依存している事態もまた、上記の抹消や攻撃、その認定をめぐる法執行環境を変動させざるをえない。この歪な解釈環境と均等を逸した法執行環境自体が持つ歴史的に特異的な時点

や状況を含めてこそ、対抗的アプローチとしての「スコア記録」の実践と新たな規制の展望が開かれるはずである。別稿を期しつつ、結びに代える⁴²。

以上

研究助成

本研究は JSPS 科研費 23K12354 の助成を受けたものです。

註

¹ 総体としてのヘイトクライム・ヘイトスピーチへの規制文脈で、何らかの「スピーチ」の特性に焦点化することには問題がある。J. L. オースティン (1991) 『オースティン哲学論文集』(坂本百大 監訳) 398-408 頁で「スピーチなしの遂行」が論じられていたように、最も苛烈な現実の排除や貶めは、暴力がそうであるように、発話なしにもなされるためである。あるいは、いわゆる「指令語」(G. ドゥルーズ) による属性付与・変更がそうであるように、発話内容というよりは、ある種の力と組み合わせられた発話による関係変化にこそ、発話もたらす害を見てとることも可能だろう。その場合における規制の構築・評価にあたっては、上記の政治的・社会的な権力をも動員する「力」の分析が不可欠であり、ここではオンライン空間のバリエーションに応じた検討が不可欠となる。さらにオンライン空間の種別によっても、その発話の持つ効力がバリエーションを持つこととなることにも留意が必要である。個々のオンライン・ヘイトスピーチの持つ効力は、リアルな対面状況でなされるヘイトスピーチの効力とは異なる一方で、集積されるオンライン・ヘイトスピーチが言説の総体としてもつ効力は、リアルな対面状況でなされるそれとは別の意味で異なりうるだろう。

なお以上の指摘が、現実の排除や貶めの害を軽視したり、その害を煙に巻くためのものになることは避けねばならない。特に、特定の社会的地位や組織内の役割を担う者がなす発話など、リアル空間における発話上の権威がオンライン空間に意図的・明示的に持ち込まれる場面については、害の分析とは独立に、その役割に基づく責任が生じることは明らかである。ヘレン・ソング (2023) 『人種差別の習慣：人種化された身体現象学』では、①住み込みとしての習慣 (34 - 36 頁)：社会的・歴史的に位置付けられた差別実践の身体的次元と、自己と関わりを持つ人種差別・人種化の経験的次元の双方を踏まえた分析 (45 頁) がなされており、参考になる。

² 本稿での検討が現実社会においてもつインプリケーションを、以下の通り取りまとめておく。本章校正中の 2024 年 1 月現在における欧州各国における親パレスチナデモの法的な取締まりや、攻撃・虐殺を正当化するための偽情報 (あるいは攻撃対象についての不確かな情報等) の拡散事例を念頭におけば、「除去手段としての規制とその濫用のおそれに伴う社会的コスト」を検討する必要性は、実践的にも明らかである。一見したところの差別的発話を除去するだけでは、差別的慣行やその基盤的規範は維持されてしまうことを、上記の自体は示している。本稿が以下で検討していくように、対抗言論が現になされることで議論が開始され、議論によるスコア (記録) が蓄積されることがないところでは、より正確な根拠に基づく主張の蓄積も構造的な平等の実現も継続しえない。ここから、情報の発信・拡散 (流通)・蓄積の「ゲート」をなすソーシャル・メディア・プラットフォームが介入することが期待される対象も、第一義的にはこの適切なスコア (記録) の実現に資する情報環境の改善にこそ存するだろう。

特に 2024 年 1 月現在においても、イスラエルによるガザへの攻撃が継続している点に、これを見ることができる。その圧倒的な武力行使により国際法違反 (ジェノサイド罪) も問われているこの状況の中、また軍人・民間人の別を問わない虐殺を正当化する数々の発話が高官において飛び交う中であって (例えばイスラエル国防大臣による「パレスチナ・ハマスの支持者は、たとえ武器を持たずスーツを着ている人であっても全員死ぬべきだ」発言など)、同時に反ユダヤ言説に至りうるデモ等がフラン

ス等で禁じられ、また米国ハーバード大の学長が反ユダヤ的であるという理由で辞職を迫られるなど、むしろヘイトスピーチ規制それ自身が政治的立場に応じて不均衡に用いられている状況こそ、総体としてのヘイトクライム・ヘイトスピーチへの規制文脈では問題とする必要がある。また戦争映像が残虐であるという理由で動画共有プラットフォームから削除されることによる戦争自体からの遠ざけの問題はロシア・ウクライナ戦争においてよく知られることとなったが、現下のイスラエルによる攻撃・虐殺においても偽情報に基づく相手方への「非人間化」と情動を喚起するための情報戦は加速している。これは所謂「言語的ハイジャック」（後述する Anderson(2020)を参照）と同様に、ヘイトスピーチの認定が恣意化されることで、重要なクライムを含む harmful counter-speech をも規制対象としてしまう問題と相通的である。ヘイトスピーチの認定と、それに基づく指弾や「キャンセル」が、対象者についてもその程度・態様についても恣意的に運用されえ、また偶然的・状況依存的なものとして事実上は通用してしまいうる（そしてそのために動員可能な物的資源やさらには解釈資源まで不均等である）ことから、むしろ「ヘイトスピーチである」という糾弾にさらされた際の対抗表現の特質をよく検討する必要がある。

³ もちろん、この時点を確認上の困難や政策的考慮の下で前倒しするものとして、抽象的危険犯がある。例えば、言葉を用いた脅迫が成立する場合、結果として現実には恐怖を生じさせなくとも、その危険において罰される。脅迫を含む犯罪類型（恐喝罪、強要罪など）も同様であり、結果として現実の恐怖や金員等交付がなくとも、その未遂が罰されることになる。侮蔑、名誉毀損、脅迫等はいずれも個人的な法益としての危機感の除去にその可罰根拠を持つ。ただし、これらはいくまでも発話が引き起こす（cause）危険の創出時点を前倒しする規定であり、後述するように発話はその瞬間において成している（do）害悪の捉え方とは異なる。

⁴ オンラインヘイトスピーチを取り巻くより広いオンライン侮蔑の現象に目を転じると、オンライン空間固有の性質に着目して、別立ての規制（本邦の2021年の侮辱罪改正のみならず、欧州での改正動向としての結果的加重犯としての致死罪導入も注目に値する。）を要求する見解がある。これをヘイトスピーチと比較することは有益かもしれない。例えば深町晋也「オンラインハラスメントの刑罰的規律——侮辱罪の改正動向を踏まえて」（2021）法学セミナー 66(12)では、オンライン空間における迷惑・嫌がらせ行為の持つ特性として、①匿名性、②高度の流通性、③永続性、④（被害者からした）回避困難性があげられている。これらの性質は、対象・目的・態様においてより限定的なオンライン上のヘイトスピーチにも当てはまる。しかし下記の理由から、これら特性が、直ちに同表現による悪影響を裏付けるわけではない。

第一に、①匿名性とされる特性は、第一に望ましい表現も望ましくない表現も助長する限りにおける道具的な価値を示唆するにとどまる。事実と反する発話一般が内在的な害をもたらすのではないように、匿名性そのものが内在的な悪さを生じさせるのではない。匿名的だからこそ、自身の社会的人格・社会的属性に囚われることなく発話ができるというのは、社会的人格・社会的属性が付与されていたからこそ抑えられていた「不適切」な発話の抑制を効かなくさせるリスクを生じさせるとともに、市民社会において重要な価値である立場にとらわれない発話を言論空間に充溢させる価値をも持つだろう。匿名性の害はこのように外在的にのみ把握されるのであり、概念の分析によって把握されるわけではない。第二に、そもそもここで議論されているオンライン上の匿名性とは、技術的に見れば仮名性に過ぎず、また実践的にも追跡可能である。いわば後の異議時点を想定した「予定的証拠」取得が原理的に存在する以上、あくまでも開示のためのコストを負担しない限り匿名コミュニケーションという約束事の下にある、という想定に過ぎない。反対に、このコスト負担が過重になることで実質的な反論を塞ぐ（モブ集団が大量に妨害的なDMを送る等）などの条件の下において、匿名性に由来する特性が規制根拠となりうる事が示唆されるに留まるだろう。あるいは、明示的であれ黙示的であれ（調べればすぐに特定可能であるという意味で）「顕名」で活動せざるをえない研究者の発言のうち、現実の所属機関における職務と強く結びつくわけではない活動領域における発話を取り上げ、組織外から所属の組織内規範を用いて「キャンセル」を働きかけることは、（組織内規範が私生活上の品位の維持や全人格的評価にまで及び、またその特定のための適正な手続きが準備されているなどの例外的な場面でない限り）上記の特性からすれば過剰な制裁と評価されるべきだろう。前稿の「キャンセル」についての箇所を参照せよ。

第二に、②高度の流通性及び③永続性とされる特性についてみよう。例えば、プライバシー侵害や「粘着」につながる場面においては、情報流通のスピードや一旦知られた後における回復の困難という帰結から見て、検索・蓄積環境に対してもたらす影響が甚大であることは疑いない。しかし、その他の表現については中立的である。表現の流通一般についてみれば、オンライン上ではむしろ対抗言論を含む発話群もまた環境的には等しく流通・永続する効果を持つ以上、集団的なヘイトデモなどの対面状況とは異なる環境およびその効果をもつ。リアル空間におけるヘイトデモは実行側の都合でなされ、カウンターデモはその探索・認識・組成・実施に至るコストを非対称な形で強いられる一方で、オンライン空間におけるカウンター発話（及びカウンター発話を支持する発話群）におけるコストは対称的である。

最後に、④回避困難性についても、媒体特性上、当該表現への一瞥の後に「見ないで済ませる」ことも通常は容易であり、ま

た対面に比すれば（双方向のコミュニケーションを遮断し）自らのイニシアティブで当該表現に多面的に向き合うことも（あるいは全く向き合わないことも）一般的には可能である。通学のための公共交通機関や大学キャンパス内でなされるヘイトスピーチは（大学に通学するとの意思と、公平な研究・教育価値の受領を毀損するために）通学している以上は回避困難だが、送り付けられた個別の意見を見たくなければ、いわばゴミ箱に捨てるかブロックすることができる。

このように、オンライン上の特性に由来する害の限定性と害の拡張性の両義性を踏まえた分析をなすものとして、Michael R. Barnes (2023) “Who Do You Speak For? And How? Online Abuse as Collective Subordinating Speech Acts”, *Journal of Ethics and Social Philosophy* 25 (2), 251–281 も参照せよ。

⁵ M. McGowan(2019) *Just Words: On Speech and Hidden Harm*, Oxford の第 2 章および同書についてのシンポジウムを踏まえた McGowan (2022) “Response to Critics” *Australasian Philosophical Review* 5 (2): 211-220 に詳しい。

⁶ この点について、スコア変動の原因の全てが「受け入れ」にあるかどうかは問わない。なお、後述するように、少なくともオンライン空間においては発語内行為の文脈依存性から、オースティンの発語行為・発語内行為についての（サールではなく）ストローソン流の解釈をとった上で発語内行為の脆弱性を認める方が有益であろう。

⁷ この拡張への批判として、Caroline West(2022), *Does Public Racist Speech Constitute Hostile Discrimination? Comments on McGowan*, *Australasian Philosophical Review* 5/2: 179-88. 特に p.183 を参照せよ。West の批判への再反論として、McGowan(2022)はその第 3 節 [The Way Sneaky Norm Enactment Generalizes] で、会話における受け入れではなく、会話全体を統制する g 規範によって卑劣な規範が指定されると主張する。しかし、これはほとんど基底となる g 規範が（会話の真偽や妥当性の変化に貢献しないスコア変更であっても、貢献するスコア変更であっても）個別的状況ごとの s 規範を設定すると主張するにとどまり、その「望ましくなさ」の主張にあたってはほぼ無内容である。さらにこの見解は s 規範と g 規範の区分を無効化させており、理論的資源として活用できるところを失ってしまうだろう。後述する Waldron (2022) はこれを規範設定の遍在性に由来する「トートロジー」と呼ぶとともに、また s 規範設定と g 規範の適用を McGowan が曖昧する点について、同様の批判を行っている。

⁸ マクガワンの提案を発展させる形で、R. Simpson (2022) “The Conversational Character of Oppression” 及び同氏の 2023 年 来日時の報告 “Oppressive Speech and Attunement” は抑圧現象について論じている。

⁹ この点を Simpson(2023) は引き継いで「優しいいじめ」現象として論じている。発話が前提などの調整を要する度ごとに、あるいはそのコストの非対称を強める度ごとに、s 規範設定の害は構成されるのである。

¹⁰ McGowan (2022) のいう「公共空間の社会的相互作用を支配する g 規範を誘発することによって s 規範を設定する」と、そのウォルドロンへの応答 (2022) を参照せよ。なお、Simpson(2023) も併せて参照せよ。

¹¹ Max Lepoutre の近時の一連の著作がこの点を取り扱っている。特に M. Lepoutre (2019) “Can ‘More Speech’ Counter Ignorant Speech?” *Journal of Ethics and Social Philosophy*, 16. (3) 及び M. Lepoutre (2023) “Hateful Counterspeech”, *Ethical Theory and Moral Practice* 26, 533–554 を参照せよ。また永石尚也 (2024) 「オンライン上のヘイトスピーチと法的介入のグラデーション ソーシャル・メディア・プラットフォームとの協働から」『ヘイトスピーチの何が問題なのか』（法政大学出版局）において論じたように、悪しき規範設定メカニズムの解明は、規範設定行為の悪性ゆえの排除を直接には正当化しない。悪しき s 規範に対して g 規範で対抗することもできるし、s 規範設定の立て直しプロセスによって対抗することもできる。この点について、第二節で後述する。

¹² McGowan(2022)の第 3 節 [卑劣な規範の設定が一般化する方法] を参照せよ。例えばムスリム排除的な発話を考えてみればよい。当該発話自体が会話における敵対性の増加と証言信用度の低下への「おそれ」を生じさせ、「自らはムスリムだがそうは思わない」、「自分はムスリムではないが、そうは思わない」と信条を明らかにしつつ反論することを、自己開示の負担と安全性の懸念から躊躇させるかもしれない。結果として、その躊躇により取り巻きと共に暗黙のうちに卑劣な s 規範の設定を行うことになるかもしれない。

あるいは、明示的に宗教的信条を用いたカテゴリーによって実際には特定の出自を持つ集団を名指す語彙を仮構し、思考の錯覚を引き起こしている点に問題を見ることがもできる。前稿でも参照したハーマン・カペレン他『パッドランゲージ』第 4 章第 3 節では「人種」の概念が生物学的事実と社会的事実を混同させる語彙として提示されるが、「信条」の概念も社会的事実内部における類似の混同を引き起こす語彙として用いられる。すなわち、会話における記述上の真偽を与えるスコアを、宗教的信条という（それまでは会話との関連性の薄い）要素に焦点化した形へと変更させることになる点にも、問題を見てとることができるかもしれない。

¹³ なお、McGowan(2019) が典型的なものとして提示する例であり、しばしば議論を呼ぶ例として、裁判例 *Bond v. Michael's*

Family Restaurant(1994)を元に考案された仮想ケースがある。店主を含む店内の会話における黒人蔑視表現が問題となった仮想ケースから、McGowanはs規範設定がもたらす問題を定式化している。具体的には、あるレストランにおいて(店主でも店員でもない)白人女性客が、別の客であるアフリカ系アメリカ人家族に対して、望みの席に座られたことの腹いせに、「あれは私たちの席だ」、「家に帰れ、もしも家があるならね」、「福祉依存でどうしようもない人間のクズ野郎」と言い放つ事例である。マクガワンによれば、このケースの発話は次のように分析される。すなわち、当該会話において許された発話や真となる発話を措定するs規範が、発話ごとに生成され、続く会話を統制する。例えば対面状況でなされた黒人排除を企図したヘイトスピーチは、相手や周囲がそれを撥ねついたり、抗弁したりする前の発話の瞬間において、黒人排除的なs規範の設定をなす、といった具合である。

¹⁴ D.Lewis(1979)“Scorekeeping in Language games”を参照せよ。特に構成的規則と統制的規則がダイナミックに変動する状況を記述するアイデアとして「スコア記録(scorekeeping)」が導入されていることが、本稿におけるMcGowanとの関係では重要となる。なぜなら、McGowanが会話における「受け入れ」を(例外は認めつつも)争い難い許容可能な事実(permissibility)の構築過程として記述しているのに対し、Lewisの元の論文においては会話における「受け入れ」はスコア運動の進化的特性の現れに過ぎず、許容可能な事実が設定されるか否かは会話にとっては(その運動の外にある参加者相互の力関係やゲーム自体の性質に依存するために)外在的なものとして捉えているためである。

¹⁵ これらはいずれもD.ルイスが挙げた例である。例えば、「平ら」を含む曖昧な文について考えてみる。「平ら」は程度や何らかの適用場面を想定することで初めてその精度が共有知識(公然の信念)に組み入れられ、使用可能となる語であり、文脈措定なしには真偽の判断はできない。例えば日常的に「地面が大体平らだ」という主張がなされた場合、そこでは物理学的に完全な平面が想定されたり、逆に陰謀論の例によく上がる地球平面説が主張されたりしているのではなく、単にコップが傾かない机を作る場面であるとか、道路の水捌け用の傾斜設計場面といった、何らかの具体的目標の下で「平ら」の語が約定され、用いられているに過ぎないだろう。

¹⁶ D. Lewis(1986), *Philosophical Papers: Volume I* 所収のDavid Lewis(1978), “Truth in Fiction”(邦訳として、樋口えり子訳「フィクションにおける真理」現代思想 vol. 234(1995年))及びその補遺David Lewis(1986), “Postscripts to ‘Truth in Fiction’”を参照せよ。特に後者では、あるフィクション作品における真理について、フィクション全体における各命題の真偽をその積集合(intersection)としての無矛盾性に基づき評価するのではなく、和集合(union)としての評価として(明示的に矛盾する個別的命題の連言(conjunction)を除外しつつ)各フィクション断片中における真理として受け入れる余地を認めることで、フィクション作品全体で言われていることを失うことなく解釈過程に取り込む提案がなされている。本文で紹介したスコアキーピング論における計画の例で見られる、反事実的条件法使用とその留保を含みながら進行する会話との並行性が見て取れる。

¹⁷ おそらくこの理由は、McGowanがJust Wordsの第2章においてルイスのスコアキーピング論を参照する際、もっぱら[例1: presupposition]及び[例3: definite description]の指示語のケースを引き合いに出し、会話における前提形成に焦点化したことが原因である。上述の通りオンラインヘイトスピーチについては、むしろ「計画」がその性質を近しくするものと思われるものの、その他の言語ゲームにおけるスコアキーピングの動きにはそもそも触れられていない。

¹⁸ D. Lewis(1979)“Scorekeeping in Language games”の[例8: planning]の箇所(p.357)における「計画の前提への組み入れは阻却可能である」ことへの明示的な言及箇所を参照せよ。この文の冒頭にある「通例通り」という確認の通り、これは計画プロセス固有の問題ではなく、スコア記録一般に通じる特質である。併せて1978年のD.ルイスの論文「フィクションにおける真理」(樋口えり子訳、現代思想 vol. 234, 1995年)を参照せよ。そもそも、このスコアキーピングの着想はそれに先だつ『コンヴェンション』(1969年)において提案していた共通知識、より正確には「公然の信念 overt belief」というアイデアを引き継ぐものである。しかしマクガワンはこの内的な連関をおそらく意図的に無視している。共通知識(公然の信念)というアイデアは、コーディネーション問題に回答する過程において提示されていたものであり、「～である」という共通知識は次の事態Aが成立している場合かつその場合に限り成り立つとされる。すなわち、(1) 集団Pの中の全員が、事態Aが成立していることを信じる理由を持っている、(2) 事態Aは、集団Pの中の全員に対して、集団Pの中の全員が事態Aが成立していることを信じる理由を持っていることをインディケートしている、(3) 事態Aは、集団Pの中の全員に対して～であることをインディケートしている、という場合である。これら条件に基づきPに属するものたちは、背景情報を相互的に帰属することによって、他のメンバーが何を信じ、どのような高階の期待を形成するのかを確定させることができる。事態Aは共通知識(公然の信念)の基底(basis)である。

差別的な発話がなされた際のスコア記録上の問題も、このPとAを引き継いでいる。仮に共通知識の条件(1)が成り立っていないならば、そもそも差別発言は差別として機能しない(それは悪態等と区別がつかない)。あるいは条件(1)が成り立っていても、

そこで成り立っている信念内容が、差別的なものなのか反差別的なものなのかは、集団内の各人において認識不能である。従って、差別的発話に関する共通知識の条件(2)の成立は、少なくともオンライン環境を前提とする限りにおいて、極めて危ういものとなる。

さて、マクガワンもおそらくこの点に気づいており、ルイスを集中的に論じた箇所では、それゆえに信念に焦点を当てた共通基盤ではなく、心理的要素を超えた外的要素も取り込めるスコアキーピングの枠組みを用いる、としていた。しかし、上述のように共通知識とスコアキーピングは対立的というよりは相互補完的である。

¹⁹ 意味論とメタ意味論の区別については、S. Kripke(1980), Naming and Necessity の p.59 (邦訳として、八木沢敬・野家啓一訳『名指しと必然性——様相の形而上学と心身問題』(1985年)68頁)を参照せよ。

²⁰ D. Ball (2018) Lewisian Scorekeeping and the Future, Croatian Journal of Philosophy 18 (54) 375-384 及び D. Ball & T. T. Huvenes (2022) "A Puzzle about Accommodation and Truth", Philosophical Studies, 179 (3), 759-776を参照せよ。Ballは本文に上げた自身の立場を Backwards-Looking Meta-Contextualism と呼ぶ。(なお、ブラングムの推論主義と時間的外在主義との関連について言及するものとして、Takaaki Matusi(2021), "Inferentialism and semantic externalism: a neglected debate between Sellars and Putnam" British Journal for the History of Philosophy 29(1) 126-145を参照せよ。

この点と関連する箇所として、McGowan(2019), Just Wordの脚注の中において、「技術的には」と前置きしつつ、(マクガワンは)スコア変動がまずは即時になされた上で、事後的に訂正や疑義によって元に戻されうる(changes the score back in the relevant respects.)と論じていた。確かに、その場限りの会話における「正しさ」を構築する種類の言語ゲームにおけるスコア変動は即時的であることが多いだろうが、他方で、計画ゲームにおいて典型的にそうであるように、スコア変動は条件付きのまま保留されることもあれば、将来における事実によってスコア変動が事後的に確定することもしばしばである。McGowanのいう即時性は、典型的には指示語ゲームなど、聞き手の意味論的理解が専ら文脈上の整合性に依存している限られた言語ゲームにおいてのみ成り立つ。この点においてMcGowanの整理は過度の一般化の誤りを犯している。ルイスがスコア変動における進化的特性を論じていたことは、修辞的なもの以上の重要性を持つ。即時性と訂正困難性を強調するマクガワンのスコアキーピング理解は、ルイスに違背するものである。進化的特性が実現するためには、それ自体の柔軟性に加え事後における「可塑性」の条件を要するものと思われる。しかし、マクガワンがs規範の事後の変更困難性を強調する段においては、むしろこの進化的特性は不可逆なものとして、いわば魔術的に措定されるかのように戯画化されている。

²¹ この点について、定型的な発達に変数となり典型とされる発話内行為が不発に終わるパターンや、その不発についての責任帰属が受領者側に転嫁される例を検討してもよい。田中優子・神尾 陽子(2007)「自閉症における語用論研究」心理学評論 50 (1), 54-63を参照せよ。また、前稿でも参照したハーマン・カペレン他『バッドランゲージ』の第9章第2節をも参照せよ。

²² この点をJ. Waldron (2021) "Commentary on Mary Kate McGowan's 'Just Words: On Speech and Hidden Harm: An Overview and an Application'" Australasian Philosophical Review Volume 5, 170-178 が指摘している。つまり、McGowan(2019)が挙げた白人客のケースでもヘイトスピーチがもたらすs規範の害が、発話的に訂正困難である事情は認められないというのである。すなわち、レストランで罵倒した上記白人客は、同じ姿をSNSで晒されることでパッシングの対象となることが予想される。その際、罵倒はアフリカ系アメリカ人への抑圧のs規範を暫定的に設定すると共に、すぐ後に続くだろう対抗的な非難によってその効力を減じられることが予想され、また当該対抗的な避難は罵倒する騙れる属性への沈黙化に向けたs規範を設定しもするだろう。単一のs規範設定がなされ、維持される場面というのは、背景となる社会的・歴史的状況を呼び起こす形で、(通勤電車や職場や授賞式等を典型として)離脱困難かつ閉じられた会話の空間において、抑止困難な形で同調する圧力に支えられた、ごく限られた場面である。

なお、例えば信用度を下げるといった一般的には害として把握されるs規範設定の営為であっても、その規範設定の上で事後に立ち向かうことが期待される場面も容易に想像できよう。例えば、立法府において首相が自身の任期中の政治的成果についての因果的な寄与について厳しい質問を投げかけられる場面を想定せよ。前稿でも参照したハーマン・カペレン他『バッドランゲージ』第2章第5節では、政治家同士が、複雑な因果的な連鎖を持つ金融危機の主たる原因が相手方とった行動にあることをめぐって舌戦を交わし、互いの二枚舌をもとに証言の信用性を下げる働きを聴衆に対して(ひいては政治的行動を変更させるべく)投げかける例を挙げている。その他、裁判において証言者の認識の整合性に対する厳しい質問が投げかけられる場面なども想定してみても良いだろう。個人ではなく集団の属性に基づく場合であってさえも、こうした信用度の上げ下げを生じさせるやり取りが許容されるべき場合はある。

²³ 想定されうる全ての危害を事前除去しようとすることは、コミュニケーションへの参加を拒否することと同値でもある。例えば当該社会で通用する「標準語」に十分馴染んでいない者は、能力的に発話機会が阻害され、会話の規範に従うことに困難を

抱えるだろうが、この発話 (locutionary) 上の沈黙化を法規制によって修正しようとするものはおそらく少数であろう。発話 (locutionary) 上の困難をケアする場面として、たとえば吃音を持つ者を含めた会話における会話スタイルの改善などが想定されるだろうが、その多くは規制によって達成されることが想定されたものではないだろう。より卑近な例として、例えば「心理的安全性」を欠いた企業における上司部下関係が自由闊達な意見交換を妨げることはありうるかもしれないが、それは少なくとも現在における法的な危害の捉え方とは距離を持つだろう。

²⁴ 引用箇所は、ジェレミー・ウォルドロン『ヘイトスピーチという危害』による。オンラインとの比較については、J. Waldron. (2021) "Commentary on Mary Kate McGowan's 'Just Words: On Speech and Hidden Harm: An Overview and an Application'" Australasian Philosophical Review Volume 5, 170-178. を参照せよ。そこで Waldron は、リアル空間におけるヘイトスピーチに対してさえも、例えばスマートフォンでの撮影・拡散を通じてオンライン空間と接続されることで、リアル空間における権力関係が、ヘイトスピーチを成す側を弱体化させる形で変動していると指摘している。他方で、こうした匿名的で距離を持ったオンライン空間における、個々のマイノリティ成員が持つ生の多様性が、オンライン上で破られる場面とは何か。典型としては、遮断が容易ではない「粘着」のケースや、多量の表現が連続してまたは共同でなされることで、オンライン上の通常の活動に支障を来たすケースが挙げられる。ただし、これらは広義のオンラインハラスメントや偽情報対策については妥当しても、オンライン上のヘイトスピーチとしての機能を果たすことは稀であるようにも思われる。

²⁵ 鳥海不二夫・山本龍彦による共同提言「健全な言論プラットフォームに向けて — デジタル・ダイエット宣言 ver.1.0」は、よど号ハイジャック記事抹消事件判決 (最大判昭和 58 年 6 月 22 日民集 37 卷 5 号 793 頁) 及びレベタ事件判決 (最大判平成元年 3 月 8 日民集 43 卷 2 号 89 頁) における「情報等に接し、これを撰取する自由」の内に、自律的な情報撰取の自由の契機を見ている。関連して偽情報対策の文脈に即して、「フィルターバブル」や集団極化が論じられることもある。しかし災害時や選挙時など偽情報が特に問題となるケースを区分けする必要がある、また「知る権利」の観点からバナーナリスト的な介入の妥当性は、事例ごとに独立に論じられるだろう。総務省「プラットフォームサービスに関する研究会」における最終報告書 (2020 年 2 月) を参照せよ。

²⁶ ちょうど 2024 年 1 月に、プロバイダ責任制限法の改正案の提出についての報道がなされた。内容としては、投稿削除などのプラットフォーム側の措置の強化と、その手続の透明化・簡素化、通知等の情報提供、さらには措置にあたっての基準の設定とその公表を求めるなどである。読売新聞 (2024/01/12) 「ネット上の誹謗中傷は迅速削除、SNS 大手に義務付けへ…法改正で削除基準の透明化も」を参照せよ。なお、本文に述べたアーキテクチャおよび主体への依存性は、より一般的に言えばローレンス・レッシングのいうコードの問題であり、当該コードへ向けられた法規制によって間接的に人々の振る舞いを制約する間接規制の問題である。ローレンス・レッシング『CODE VERSION2.0』(2006) 第 5 章「コードを規制する」及び第 7 章「何が何を規制するか」を参照せよ。また成原慧 (2021) 「媒介者責任の再検討 プロバイダ責任制限法改正及び関連する取り組みの意義と課題」法学セミナー No.803 及び成原慧 (2021) 「インターネット上のヘイトスピーチとその規制」『ヘイトスピーチ規制の最前線と法理の考察』(2021) も参照せよ。なお、規制の手法選択を検討するに際しては、関連して総務省「インターネット上の誹謗中傷をめぐる法的問題に関する有識者検討会取りまとめ」(令和四年五月) も参照せよ。

²⁷ さて、このようなスコアボード構築プロセスを通じた s 規範の通時的確定プロセス性とその保障 (assurance) の上で、g 規範の組成及び g 規範による s 規範への重みづけ (制約) もまた重要なものとして現れる。この点については、国家の役割が、直接的あるいは DPF を通じた間接的な形で迫り出すことになるだろう。たとえば直接的な形では、政府言論としての反差別の保障が典型であり、間接的な形では偽情報を含む私的な過剰認定やその加熱現象を議論に開くためのフォーラム提供のための共同規制等が許容される。この点につき、キャンセル・カルチャーに代表されるように「公共」を標榜した市場への間接的影響を放縦化させず、その効力をよき社会の構想に基づき現実社会に位置づける観点から、DPF がゲートキーパーとして求められる役割もまた限定づけられる。例えば、ウィッスルブLOWER (通報) モデルによる社会的圧力を促進・支援する形態ではなく、継続的監視と改善のプロセスを経るゲートキーパー (chaperone) モデルに基づき、より確からしい根拠の蓄積・参照に裏付けられた議論環境の構築に寄与する介入形態の模索が求められるだろう。

²⁸ フランスにおけるいわゆる「インターネット上のヘイトコンテンツ対策法」における大半の条項が違憲とされた上で公布・施行された経過において、これを見ることができる。

²⁹ R. Langton (2018) "Blocking as Counter Speech," in New Works on Speech Acts, Oxford, 144-164. を参照せよ。また関連して、R. Langton (2012) "Beyond Belief: Pragmatics in Hate Speech and Pornography," In Speech and Harm: Controversies Over Free Speech, Oxford 72-93, R. Langton, (2018) "The Authority of Hate Speech," in Oxford Studies in Philosophy of Law (Volume 3), Oxford, 123-152. も参考になる。

- ³⁰ Laura Caponetto の近時の複数の著作による。本稿執筆にあたっては特に、Caponetto, L. (2018) “Undoing Things with Words” *Synthese*, 197, 2399-2414, Caponetto, L. & Cepollaro, B. (2022) “Bending as Counterspeech”. *Ethical Theory and Moral Practice* 26 (4):577-593, Caponetto, L. & Cepollaro, B. (2023) “The Philosophy of Counter-Language” in *Counterspeech: Multidisciplinary Perspectives on Countering Dangerous Speech*, Routledge を参照した。
- ³¹ 同上、Laura Caponetto(2019) を参照せよ。
- ³² 例えば、Wikipedia の編集が一時停止される場面、すなわち論争が加熱したテーマとそのページについての「凍結」などを例に挙げることができよう。あるいは、Twitter におけるリプライ制限なども「非決定」戦略の技術的支援として捉えることもできよう。
- ³³ 技術的な例としては、Twitter (現 X) におけるコミュニティノートにおける評価メカニズムが参考になる。個別に付されているコミュニティノートは信用できるものからそうでないものまで幅を持つ一方で、同機能についての信用性評価が信頼の置けないものであるか否かは記録されていることが公にされている。すなわち、単純な情報操作手法に耐性を持たせるべく、「過去の評価において、意見が相違することのあった協力者の間で意見が一致した場合」にステータス上の寄与を認め、反対に信頼の置けない評価者からの評価の価値を一般的に下げられていくことで無益な評価者とし、長期的には信頼の置ける評価者による評価が量的に高まることを期待される仕組みである。こうした評判メカニズムを健全に機能させる仕組みを、スコア記録の実装という観点から再記述することも可能であるものと思われる。詳細は下記を参照せよ。 <https://communitynotes.twitter.com/guide/ja/contributing/writing-and-rating-impact>
- ³⁴ この点について、デジタル立憲主義と規制主体: 国家による DPF 規制のデュープロセス、国家によるローカル規範規制のデュープロセス、DPF によるローカル規範規制のデュープロセスの関連を考察することが求められよう。
- ³⁵ 上述した語彙を引き継げば、これをメタ意味論的な正義の要求として定式化することもできるかもしれない。この着想の元として、D. Ball(2020), *Metasemantic ethics*, *Ratio* 33(4) における意味固定をなす者が果たしうるメタ意味論的誤謬のパターンの整理が参考になる。とりわけ、C. サンステインが定式化した「不完全に理論化された合意 *incompletely theorized agreement*」の問題との比較は、その主体や場面との比較の観点から有効であるものと思われる。
- ³⁶ ややラフな言い方ではあるものの、例えばある種の有害と判断されうる発話に対して DPF 側が警告表示を出したのに対して、発話者があえて警告を無視して修正なしに投稿した場合には、その発話上の意図性を推認させ、またそれを蓄積していくことでスコアボード記録上の「誤審」を少なくする対応などが考えられよう。本文に述べたように、発話者に理由を聞いただし、それに発話者が誠実に応えるといったゲームは所与として成立しているものではない。ただし、スコアボード構築・管理が機能する限りにおいては (あるいは DPF が積極的にこのスコアボードの構築に従事する限りにおいては)、ある種の「言質」を収集・蓄積し、議論に先立つ予防的な証拠として参加者にアクセス可能にするゲームとしてオンライン空間上の熟議を組成することは不可能ではないものと思われる。
- ³⁷ Judith Butler(2022), “Categories by which we try to live” *European Journal of Philosophy* 31 (1):283-288
- ³⁸ R. Langton(2012) を見よ。なおこの点は、ターゲット集団に属する者の正統な権限を剥奪する文を考えてみればわかる。例えば、シオニズム運動の祖たるテオドール・ヘルツルの「聖地には住人がいないのだから、土地を持たぬ民族たるユダヤ人は罪悪感なしに占領できる」という文を例に挙げよう。この文には文面上、特定の人々に対して明示的にその排除やランクづけを変更する悪性はないものの、そこに現に住まうパレスチナの人々そのものを無化している。あるいは現代においても、同様の発話の例には枚挙にいとまがない。例えば Judith Butler のインタビュー “Palestinian Lives Matter Too: Jewish Scholar Judith Butler Condemns Israel’s “Genocide” in Gaza” (2023) では、ガザへの電力供給の停止に伴い死者が出る点について質問を受けた際のイスラエル前首相ナフタリ・ベネットの回答、“Are you seriously talking about Palestinian civilians?” が紹介されているが、これもまた「(死者が出るとしてパレスチナには) 市民に値するものはない」あるいは「市民がいたとして話すに値しない」という前提を導入するものである。このように上記の文は、特定の人種の無視あるいは殲滅を当然視する前提を導入するものでありつつ、特定の属性を名指すことがなく、むしろ名指すことがないことによってこそ潜在的な「犬笛」効果あるいはマニピュレーションを及ぼす。この点において、より一層の悪性を有する。それにもかかわらず、上記の文章はヘイトスピーチとして認識される典型からは外れ、単なる誤った文として一般的には認識されるに留まるだろう。
- ここでの「マニピュレーション」の語は、三木那由多 (2022) 『会話を哲学する』による。三木那由多は『会話を哲学する』第 1 章 [コミュニケーションとマニピュレーションの関係] の中で、会話当事者間において (それを本心として信じるか否かは別として) 「知らないふり」をすることが困難となる公共的な約束事としての会話の前提や文脈を構築する側面を強く持つものを「コミュニケーション」と呼び、このコミュニケーションを通じて互いに相手の心理や行動を、自分の望む方向へと変化させようとする

側面を強く持つものを「マニピュレーション」と呼ぶ。例えば、先ほど挙げた「聖地には住人がいないのだから、土地を持たぬ民族たるユダヤ人は罪悪感なしに占領できる」という文が発話された時には、文字通りに「聖地には人と呼べるような住人はいない」あるいは「聖地には不当な居住者しかいない」という共通の信念を構築することが、コミュニケーションの対象となっていたかもしれない。他方で、明示的なコミュニケーションを回避する形で、他者に対しては「以上を通じて、聖地を自らのものとせよ（障害となるパレスチナの人々を排除せよ）」という指令や促しが含まれている、とすることは妥当であろう。このように、ある発話の意図やコミュニケーションを決定するというよりも、他の現実上の力学との関係において現に成し遂げている共同関係こそが、発話の害の分析においては焦点となる。

もちろん、最も低いランクづけの極北として、この例を位置付けることもできる。和泉悠『悪い言語哲学入門』第8章[ヘイトスピーチ]でもその発話の悪さをランクづけ機能に基礎付けている。しかし和泉が、本文に上げたような文、すなわち真偽が問いえる一方で、その真偽を評価によって糊塗する発話あるいは真偽について的事实を作り出す実力に裏打ちされた発話について、その悪性をどのように見出すのかは明らかではない。この点については、フランスにおいて「共和国の価値」への同化を求められるとともに、(アルジェリア戦争を隔てた) 世代ごとに異なる記憶を持ち、分断されてもいるフランス内のアルジェリア系の人々を想起することができよう。ミシェル・ヴィヴィオルカ『差異』の第8章(特に[集合的記憶の形成からその公的な表現へ]の箇所(237-241頁)を参照せよ。

³⁹ 同上、三木(2022)を参照。上記の混乱の原因は、発話の意味を確定するプロセスとしての「コミュニケーション」と「マニピュレーション」が、発話者の態度に収束するでも共同体の承認に収束するでもなく、その間で浮動(振動)し、正当化された発話の地位を争っている事態に由来する。すなわち、発話側が「マニピュレーション」の意図を詳らかにせず、他方で「コミュニケーション」で正当化される形式と内容において表示された事柄以上のことをなす際に生じる問題であるなお、このマニピュレーションの責任を問う仕方につき、三木はコミュニケーションのレベルでの責任とは区別された、もたらされる結果やその予見についての「より一般的な行為の善悪の次元」での責任を要求する。この応答責任の実践は妥当であろうが、他方で、本文で述べるようにこの責任は規制の正当化根拠にはならない。

⁴⁰ Derek Anderson(2020) "Linguistic Hijacking" Feminist Philosophy Quarterly6(3)のDPF上での実装として理解できる。なおAndersonは、上述したBall(2020)を引きながら各種の倫理的義務を導いている。併せて参照せよ。

⁴¹ 上記のJudith Butlerのインタビュー(2023)による本人の発言による。引用箇所は拙訳であるが、全文の翻訳は下記にて読むことができる。<https://note.com/bashir/n/n78fb1d686563>

併せてジュディス・バトラー(2019)『分かれ道』(大橋洋一・岸まどか訳)、とりわけ第7章における(ホロコースト)の語り時間が時間を越えて伝達される際の修辞利用の不可避性と、その修辞が(指示対象となる出来事をアーカイブする目的の下で)出来事から擬似的に独立の(その意味で危険な)「結晶化」に帰結してしまうことのジレンマ状況への言及箇所を参照せよ(360-386頁)。特に後者の結晶化に見られるように、集団的な語りは「言語が記録し、保存し、伝達する素材を不可避的に加工もする」。このことから例えばホロコーストをめぐる「言説的奪取」を生み出す危険を持ち、「決して忘れるな」を新たな記憶抹消と土地収奪・軍事行動の政治的正当化のために用いてきた、とバトラーは述べている。

⁴² この点につき、現下のプラットフォーム規制の状況に即して具体化するものとして、情報法にかかる論集に寄稿を予定している永石尚也(2024 予定)「リスク管理主体としてのプラットフォーム事業者」が刊行予定である。刊行の暁には併せて参照されたい。



永石 尚也 (ながいし・なおや)

[専門] 法哲学

[主たる著書・論文]

永石尚也(2024: 近刊)「オンライン上のヘイトスピーチと法的介入のグラデーション ソーシャル・メディア・プラットフォームとの協働から」『ヘイトスピーチの何が問題なのか』(法政大学出版局) 所収

永石尚也(2021)「公共空間における情報識別と警察活動についての覚え書き」国際関係研究所報 56 巻

永石尚也(2020)「プライバシー・監視・アーキテクチャ 「AIと法」の余白」『法政策学の試み(法政策研究第20集)』(信山社) 所収

[所属] 情報学環 准教授

[所属学会] 日本法学会、応用哲学会、日本医事法学会、科学技術社会論学会等

The Regulation of Hate Speech and Inherent Limitations(2) : On the Nature of Illocutionary Norm-Enactment and Defeasibility

Naoya Nagaishi*

This paper aims to shed light on the desirable forms of regulation of online hate speech and the conditions for its 'desirability' through a linguistic analysis based on an illocutionary approach.

As noted in the previous paper, discussions on the regulation of online hate speech often assume that typical speech that would be considered hate speech if made in real space (e.g., on social networking sites, video sharing sites, bulletin boards, or in some other form) is made online, and then discuss whether it is possible to analogies or extend the regulation of hate speech in real space, and if so, under what conditions. However, this type of discussion is conceptually flawed in that it introduces the assumption that hate speech, which is supposed to be defined in terms of its real-life function of exclusion, humiliation, etc., can be identified from the type of content of speech.

Against this background, this paper undertakes the following three tasks. First, it is argued that the mechanism for setting conversational permissibility facts (the mechanism for setting situational norms) for toxic speech made online functions differently from the conversational situation. This is shown through a review of the arguments in the philosophy of language. This review distinguishes between sneaky norm enactment, which should be eliminated, and norm-setting that requires other interventions. Secondly, in the light of the characteristics of online space and online communication as seen in the previous paper, the mechanisms that give rise to cases of misfire, abuse, and misuse of such normative settings, which deviate from the target population, will be clarified. From this, to deal with the above-mentioned cases, in addition to simple 'objection' as a type of counter speech, 'blocking' to make implicit assumptions visible, 'bending' to modify the assumptions of speech towards a more egalitarian one, and 'non-determination' to limit the efficacy of speech are presented as intervention methods to be compared and and review. Third, based on these considerations, a jurisprudential examination of the actors (digital platforms or governments) that perform interventions that reduce the efficacy of sneaky norm enactment and promote egalitarian norm enactment in online situations, and the conditions for permissible modes of intervention.

* Interfaculty Initiative in Information Studies, The University of Tokyo

Key Words : Hate Speech, Speech Act, Content Moderation, Counter Speech, Scorekeeping, Truth in Fiction.