

オンライン・ヘイトスピーチ規制の法内在的制約（1）

—そのコミュニケーション的 / 空間的特質

Regulation of Hate Speech and Inherent Limitations (1) :
On the Nature of Online Space/Communications

永石 尚也*
Naoya Nagaishi

はじめに

本稿では、オンライン上でなされるヘイトスピーチの性質とその害の特性を、夙に近時において新たな発展を遂げてきた言語哲学上の発語媒介行為・発語内行為の観点¹から明らかにした上で、法的介入の態様の望ましさを判断するための基底的考慮事由の特定を目指す。しかし、前者の手法的な部分（発語媒介行為・発語内行為による分析）については次稿（2）以降に譲り²、本稿ではいわゆる表現の脆弱性に発する議論からこの問題を分析する。すなわち、オンライン上の規制作用に晒されたヘイトスピーチ（及びそのように誤認され、また他者に誤認されうると予期された発話）の脆弱性のみならず、ヘイトスピーチに晒された対抗言論の脆弱性につき、発話上の沈黙化や抑圧等の「悪しき」効果の発生機序にかかる既存の議論をオンライン空間の諸特性に照らして分節化し、適用範囲を画する手法の開拓を目指す。

発話の害はリアル空間とオンライン空間とで

異なった発生・構成機序を持ち、主にその技術的特性から発話への回避手段・対抗手段もまた異なる性質を持つ。さらに、コンテンツモデレーションなど発話そのものへの不可視化や沈黙化もまた、国家という規制主体以外のデジタルプラットフォームによる準・規制作用を介して可能となる。この期待の下、従来のリアル空間での発話の危害及び脆弱性にかかる議論の射程は自ずから限定されるとともに、上記空間特性に応じた新たな非規制的で協動的なアプローチを開拓する必要性が浮き彫りになる。

本稿の目的は、特定の表現特性が民主主義ないしは思想の自由市場を支える価値を毀損する等の（発話にとって）外在的な価値に着目して評価する議論に先立ち、まずはオンライン上のヘイトスピーチの形式的な特質を踏まえ、発話が置かれた社会状況・技術的環境に相関的に、当該発話の害が回避・対抗困難な形で現実化する条件を明らかにすることにある。

* 東京大学大学院情報学環

キーワード：ヘイトスピーチ、尊厳、言語行為論、コンテンツモデレーション、政治の状況、アイデンティティ政治

第1節では、オンラインヘイトスピーチが集団へもたらす害について、集団的属性と個人に対する害との関係及び非対面状況におけるオンラインヘイトスピーチの言語的特性を介した害の構成・発生機序から議論の整理を施す。これにより、リアル空間におけるヘイトスピーチとの議論上重要な差異を生み出す諸論点を取り出す。第2節では、上記の害の特質を踏まえた上で、オンラインヘイトスピーチの可罰性を問う前提を明らかにするべく、害の構成・発生機

序と国家作用としての刑罰執行との間の距離から、可罰性概念を限定する法内在的な制約を整理する。ここから、法内在的な制約としての害・規制のリスクトレードオフ関係を法的介入・不介入の場面ごとに分類する。第3節では、対抗言論の脆弱性にかかる議論蓄積の中に本稿で見てきた発話内的・発話媒介的アプローチを接続する可能性とその位置付けを示し、次稿の言語哲学的分析への接続の展望を示す³。

1. ヘイトスピーチの害の種別と、規制による害の種別

1.1 害をなす発話の特徴とヘイトスピーチの固有性

1.1.1 発話のもたらす害の区分

改めて確認するまでもなく、そもそも特定の民族的・宗教的属性を理由とするヘイトクライムや差別的取扱いが、特別にそれらを括り出す個別規定を持つか否かにかかわらず、広く法的に禁じられることは疑うべくもない⁴。そうした物理的暴力や不当な取扱いに至らないヘイトスピーチも、特定人に対する侮蔑・中傷や威圧、脅迫、プライバシーや名誉を脅かす事実の流布、暴力の扇動等の行為類型に重なる場合には規制を免れない。しかし、こうした行為類型の近傍に属する具体的発話群のうち何をヘイトスピーチに属するものとし、上記行為群と同様に禁じられるべきものとするかについては、理論的にも実践的にも、その害を弁別し、慎重な検討を要するべき多くの理由がある。

ここで冒頭に挙げた発話媒介行為・発話内行為による害を分けて論じることが一つの助けになる。詳細は次稿に譲るが、発話媒介的アプロ

チとは、端的に言えば、「それをいうことによつて (by) ものごとをなす」その因果的効果から、発話の特徴を把握するアプローチを指す。例えば、偽情報拡散類型は、基本的にはそれが惹起する信用毀損・業務妨害（刑法第三十五章 信用及び業務に対する罪等）や、世論形成⁵・選挙等を含む広義の政治過程に対する効果（公職選挙法 第十六章 罰則等）⁶、公正な競争（独占禁止法、不正競争防止法）や健全な金融市場の毀損（金融商品取引法第六章 有価証券の取引等に関する規制等）等、発話外在的な社会的害に依存する形で「害」を実現するのであり、単に内容が真実と異なる情報であるとの発話内在的要素だけで「害」をなすものではない⁷。逆に真実を述べることで「害」をなす場合がありうるのは、プライバシーに関わる真実を伝えることを念頭におくだけでも明らかであるだろう。扇動・教唆類型でも、基本的にはそれが惹起する非言語的な

暴力や暴動等の危険が実現化するか、差し迫ってなされる蓋然性が高い状況においてその「害」を実現する。相手にその決意を生じさせただけで「害」を作り出すというのは破防法等のごく例外に属する場合には存在しなくはないものの、理論的に大いに異論含みであり、かつ、その罪としての成立は当然に限定的である。ましてや（広範な文脈に依存する効果を検証することなしに）単にその文面的特性のみで「害」をなすとする考えは、およそ賛同を得がたい。このように上記の類型に属する発話は、その言葉によって（by）因果的に接続された「害」をなす類型であり、その拡張を（特別法によって）行うにせよ、発話とは独立の外在的な価値（害）によって支えられており、規制にあたってはこ

の価値（害）の証明が要される。

他方、より間接的に、特定の発話をなすことそれ自体で、続く発話・行為の可能性を制約する発話の側面もある。この捉え方は、いわば「それをいうことにおいて（of）ものごとをなす」文脈規定効果に着目する点から、発話内アプローチと呼ばれ、個別的場所・個別的文脈の下での発話の妥当性や信用性を設定する規範を変動させる点に「害」の現れを見る。ヘイトスピーチに属するとされる発話が持つもまた、具体的名宛先において一般的に生じる（「実体的」な）危害と、（証言の信用性・解釈等にかかわる認識的不正義や、行為発話への沈黙化・抑圧等を通じた）個別の状況的規範の変動とに区分できよう。

1.1.2 表現規制のもたらす害 正統性のバックドア問題

更なる区分として、ヘイトスピーチとその他の差別的・侮蔑的発話類型を峻別する基準の問題がある。差別的・侮蔑的発話と共にヘイトスピーチがなされることが傾向としてあるとしても、また当該発話が職場など離脱困難な場所的限定や地位関係・力関係においてなされる場合にはそれらが近接する場合がある⁸としても、ヘイトスピーチを伴わずして生じる害と、ヘイトスピーチそのものによる害とを分離する要請が働く⁹。目下、環境に即した対応が情報類型ごと、また事業者ごと¹⁰に、継続的な制度的整備とその見直しの中にある¹¹ことはこれを示しており、さらに人権、デュープロセス、説明可能性等の諸原則¹²に基づいてソフトな秩序形成を促進する「協調」の模索もまた上記の要請に対応したものである¹³。

この要請の根拠は何よりも、規制可否とその条件・態様を論じる文脈における「ヘイトスピーチ」をめぐる法の制定・解釈・適用それ自体が、他ならぬ国家による権限行使を前提とした営みであることに求められる。つまり、上記の中心事例が生じさせる害や行為類型が折り重なって生じる傾向性等を前提としつつも、当該中心事例からの場当たりの類推や傾向性による拡大処罰に至らぬよう、適用段階における手法の統制はもとより、それに先立つ解釈を恣意化させることをも避ける統治手法を統制する必要性に由来する。

これはオンライン空間においてヘイトスピーチがなされる場合にも当てはまる。一般に、ある表現類型Sのもたらす害類型Hについて新たな法規制類型Rを求めることは、当該規制

作用への制約を制度的均衡を目指して実行する装置なくしては、制定法の解釈・適用を通じて実現する権力に対して概括的に正統化を付与することと同義である。ましてやオンライン空間の場合、発話そのものに対する技術的介入（網羅的走査に加え不可視化等を含む）が可能である。そのため、類型Sにあるとされた個別表現sの選択段階で恣意性がある、あるいは少なくともその不信が生じるならば、類型Sを超えて個別表現sの生産・流通は阻まれようし（萎縮効果）、類型Hの状況に依存した具体化であるところの個別の害hの選択段階で恣意性がある、あるいはその不信が拭えなければ、リアル空間以上に萎縮を働かせる事態を招くだろう。

いわゆる表現の脆弱性に即してコミュニケーション中立性及び主題・観点・発話者の中立性が規制主体に要求される¹⁴のは上記事情によるが、オンライン空間におけるヘイトスピーチ

を規制しようとするれば、（物理的な威力・暴力行使を原理的に持たない以上）まさに文字・画像・映像等の情報内容に即して規制可否の判定がなされるより他はなく、上記の中立性要求に否応なく抵触する。ここから、抑止の必要性を満足させるための介入を限定する内容・手続上の規律と、それらへの（司法）審査がともに厳格になされるべきことに帰着する¹⁵。

以上より、他ならぬヘイトスピーチに固有の問題を他の差別的・差別助長的な行為類型と混ぜ込むことなく、取り出す必要がある。殊にオンライン上のヘイトスピーチの害を把握するためには、発話の文脈が個別的・複層的に折り重なり、多アクターが部分的にのみ空間をともにするオンライン空間の特性を踏まえ、同表現に属する（べき）発話類型をその基準とともに同定する必要がある。

1.2 ヘイトスピーチ規制をめぐる困難① 属性を介した害及びその回避・対抗手段選択の文脈鋭敏性

1.2.1 規制をめぐる多様性と不一致 規律根拠、管轄、規律対象をめぐる普遍性と個別性のずれ

さて、オンライン上のヘイトスピーチに属する発話類型について、それが生じさせる多種多様な害とその回避・対抗手段の選択可能性から同定基準を導くためには、比較対象としてリアル空間を前提とした既存のヘイトスピーチをめぐる議論蓄積と現実の規制（その拡散）状況を把握することは有益であろう。そもそもヘイトスピーチとして把握されてきた中心的な事例、例えば比類なき道徳的悪としか断じ得ないジェノサイド肯定の発話やホロコースト否定論を基礎とする一群の民族的・宗教的集団への蔑視的・排除的発話は、その歴史的経緯および社会状況

に照らし、上記1.1.1で見た二種類の「害」、即ち発話された個別的な文脈をほぼ問わずに強制的に変更させるとともに、実体的な人類全体への危害をも生じさせるものとして、社会におけるコンセンサスが築かれてきた。（刑）罰の執行を含めた法規制その他の対応も各国においてこれまでも実践として積み重ねられてきたし、民族的・宗教的集団に対する憎悪の表明や集団殺害・テロリズムを呼びかける憎悪表明が反差別の理念に抵触する点についても、国際的にも幅広い合意形成と勧告等の実践が蓄積されてきた¹⁶。

しかしより一般的に、具体的になされた個別

の発話について、「何が規制に服すべきヘイトスピーチに該当し、何がそうでないか」を識別・判断する基準については合意形成がなされておらず、またその規律も様々となっている¹⁷。よく知られた規律の例としては、ホロコーストの歴史を持つドイツ刑法における民衆扇動罪（第130条1項）があげられるだろうが、オーストラリアの1975年連邦人種差別法18C条の上導入されたオーストリア刑法（1995年）における集団に対する暴力の勧奨（第80.2A条）、その成員に対する暴力の勧奨（第80.2B条）、ジェノサイドの唱道（第80.2D条）等への刑罰付与も特徴的である。後者では「人種、宗教、国籍、国民的若しくは民族的起源又は政治的意見」によって識別される集団に対して、故意に威力・暴力の勧奨を行うことが禁じられるが、この根拠は専らオーストラリアの白豪主義という歴史的背景、即ち、文脈の個別性に依存する。一方、管轄内の規律対象が広範化・普遍化する結果、具体的な害の発生は問わず、勧奨行為の時・場所・方法も問わない。この結果、社会的状況やオンライン空間の技術的環境の変化を通じ、広範な行為に及ぶ¹⁸。

これらと比較した際に、規制の緩やかな国として挙げられるのは米国や日本である。例えば

1.2.2 難点1 属性と集団をめぐる普遍性と個別性のずれ

ただし、上記の不一致は単なるグローバルな規律形成の不徹底や怠慢ではなく、下記の理論的・実践的な難点があることによる。第一に、そもそものヘイトスピーチの定義をめぐる理論的難点がある。直観的に合意できる代表的な例をリスト的に並べたとしても必要十分な定義に

本邦におけるヘイトスピーチ解消法（本邦外出身者に対する不当な差別的言動の解消に向けた取組の推進に関する法律（2016年））では、「本邦外出身者に対する不当な差別的言動」を次の三要素、すなわち、（ア）専ら本邦の域外にある国若しくは地域の出身である者又はその子孫であって適法に居住するもの（以下この条において「本邦外出身者」という。）に対する差別的意識を助長し又は誘発する目的で、（イ）公然とその生命、身体、自由、名誉若しくは財産に危害を加える旨を告知し又は本邦外出身者を著しく侮蔑するなど、本邦の域外にある国又は地域の出身であることを理由として、（ウ）本邦外出身者を地域社会から排除することを煽動する不当な差別的言動をなす言動である、との要素から定義する。（ア）において多様なエスニックマイノリティーを対象として含まない過小包摂や、（イ）（ウ）における該当行為の文言上の曖昧さもさることながら、最も特徴的な点として、同法は制裁を含む執行発動根拠を持たないいわゆる理念法である点が挙げられる。すなわち、直接に集団侮蔑を法的な規律対象とするものではない。この点で、上記ドイツ刑法を筆頭とする諸法とは大きな距離を持つと指摘されることも多い¹⁹。

はならない上、歴史的・社会的状況に鋭敏に反応するヘイトスピーチ現象について何らかの還元的な定義を求めることは（逆に定義から逸脱する例を示す反論を即座に招くため）、規制可否という議論の焦点を逸する事になる。

その理由として、**特定の属性と「それを持つ**

とされる現実の社会集団」ととの間の距離の問題がある。「人種」や「皮膚の色」を筆頭として、過去に社会的に抑圧されてきた特定の集団的属性に関連づけずにヘイトスピーチがもたらす害を把握するのは困難である一方で、当該属性の多くは生物学的・解剖学的その他の明瞭な性質に規定された概念というよりは政治的論争の最中にある社会的構築物であり²⁰、マクロ的に変化に富む歴史的・社会的状況に相関的である。この結果、ある同じ名称で呼ばれる属性を（自己の一部として）持つとしても、それへの帰属やコミットメントの幅は激しく、同地域の同時代の内、さらには同世代の内であっても²¹、制度ごと・社会ごとにその可視化や取扱い上の差異も複層的である。

もちろん、過去時点の抑圧という一点において特定の属性を括えることはできるわけだが、その場合には当該カテゴリーは事実として属性を共有する具体的な集団を欠くことになる。また規範的にも、(たとえ日常的には通用しつつも)誤ったカテゴリーを使用する負の効果として、既存

の規範の強化及びそれによる周縁化を招く危険がつきまとう²²。

ここで、その属性は専ら認識的なものであるとして、記述的にはある時点の、ある社会における（誤認を含む）認識により外延を記述する作業による解明で十分とする見解もある²³。ただし、その作業は概念の内包的な明晰化ではないため、誤ったカテゴリーを認識上は真なるカテゴリーと取り違える危険を冒すのみならず、カテゴリー間の比較及び価値的な順位づけが困難に至るになる。この事情のために、新たなカテゴリーが提出されるごとにその包摂／排除の決定を繰り返すこととなり²⁴、また当該カテゴリーの拡大傾向ゆえに²⁵、規制態様もまた制限的なものに留めざるを得ないことに帰着する²⁶。結果、その時々において政治的に無力化された社会集団を過不足なく取り出し、そのニーズを満たすために「ヘイトスピーチ」という語の分析から発して有効に対象を区切ることは困難に至る。

1.2.3 難点2 属性と人格形成をめぐる普遍性と個別性のずれ

第二に、特定の属性と「それを介した人格形成及び人格毀損という出来事」ととの間の距離の問題がある。過去に社会的に抑圧されてきた特定の集団的属性は、ミクロ的に個人の内部において人格の一部に組み込まれ不可分となることもあるが、他方で当人の全人格は当該属性に尽きるものではなく、複数の属性の複雑な束ね上げによって構成される。すなわち、抑圧の源泉はある一つの属性に限定されない。逆にある一つの属性を持っていることが抑圧されているこ

とを意味するものでもない²⁷。集団と個人とは、ある時点におけるある属性の共有や当該属性への社会的視線によるカテゴリーにおいて包摂関係にあるとともに、その包摂関係もまた時間の中で揺らぎ、離脱や新規参入、あるいは社会的な呈示の過程で変容する。そうである以上、人格自体の複層的構成ゆえに、(法的紛争事例としての団体の内部関係を取り上げるまでもなく)一定の緊張関係に立たざるをえないだろう²⁸。

もちろん、当該属性が当人の良心や信条と

いった深く人格を規定する領域に根ざしていたり、人生選択に影響を与える変更困難な特徴(例えば心的外傷を含む不可逆的かつ深刻な被害などの出来事そのもの)を証跡する性質²⁹等であれば、事実的にも規範的にも人格と当該属性との結びつきは異論なく認められよう。しかし、そこにおいても何が人格に深く根ざしているのか、あるいは不可逆に当たるか、さらに誰がその被害を共有する者たちの範囲に属するとされるのかは、多くの場合論争的であらざるをえない³⁰。ましてそうではない属性であれば、たとえ引き剥がしがたい性質を持つとしても、少なくとも(価値多元性を標榜する現代社会においては通常)複数の集团的属性が重疊的に作用するものとして取り扱われるだろう。当該属性を取り巻く社会的階層間の移動・行動選択の可能性や意見・判断(その表明)の可能性が当

1.2.4 定義の放棄?

これらの難点から、特定の属性に照らしたヘイトスピーチの定義そのものを放棄する誘惑に駆られるかもしれない。たとえば、特定の属性に基づく差別的雰囲気やの瀾漫という事態をもたらす発話一般を「ヘイトスピーチ」とする、あるいは社会に内在する差別的構造を発見するための議論喚起力のある言葉遣いとして用いることを許容する、といった解決策である。明確な差別的な社会構造の存在があるならば、個別の発話はその発現そのものであり、被差別集団への危害の告知に類した作用を果たすことになるはずだ、との主張がこれにあたる。

しかし、第一に、そもそもどの属性が差別と関連するか、どのように発話が差別的な雰囲気を

該属性と強固に結びつくことで実現困難となるなど、抑圧的社会構造を持つ状況では別であるが、それは属性そのものの問題ではなく、属性との結びつきに由来する問題である。

翻ってオンライン空間を介した「人格」は、後述するようにその形成段階及び(選択的)呈示段階において、リアル空間よりもその場の規範にフィットする「顔(=面目)face」として操作可能性を介している側面を持ち、またその側面の利益を有している³¹。これらの結果、当該属性を介したヘイトクライムや差別的取扱いが当人に直接に届くのと質的に異なり、ある属性を介したステレオタイプの表現それ自体による当人の人格性の毀損という害は間接的にのみ当人に届くことを前提に、規制を組み立てる必要が生じる。

瀾漫させているかの認定を恣意化させることは、逆に(ある社会における特定マイノリティ集団による別の更なるマイノリティ集団への差別・抑圧の実態を告発するなど)正当な問題提起をなす発話をも沈黙化・抑圧する事態を招くこととなる。さらに、これら認定を規制にあたっての解釈の基礎とする場合、むしろ特定の発話の社会的意味を固定・強化してしまうことになる。例えば、ヘイトスピーチに属する発話がなされた際、専ら発話者の社会的評価が毀損される社会となるか、それとも発話を受けた者の社会的評価(あるいは名誉感情)が毀損される社会となるかは、現実の社会の状況に照らせば、少なくともその時点では多くの社会が不確定であらざ

るをえない³²。その不確定性にもかかわらず、規制により前者を望ましいものとして確定的に選択することは、当該発話そのものをステイグマ付与的なものに固定・強化する点において、被害を受けた者の救済にはむしろ反するだろう。

第二に、差別的構造を発見し、告発するために「ヘイトスピーチ」という語を用いるとすれば、翻ってその語自体が自分に対しても望ましくない発話や嫌悪をもよおす発話をラベルづけるための語として誤導的に使用されうることをも一般的に許容することを意味する。さらにそうした慣行が続けば、適切に発見・告発した状況においても、誤導的に使用しているのだろうと軽んじられる事態を自招する結果にもなるだろう³³。

以上の検討で明らかになるように、規制文脈

で求められているのは、ヘイトスピーチの**還元的な定義**をめぐる探求（やその放棄）というよりも、規制に服すべきヘイトスピーチとそうでないヘイトスピーチの**概念的な区別の改訂実践の適切性条件の探求**である。この改訂実践のためには、上記のように特定の属性とそれへの差別的な社会構造から発してヘイトスピーチの対象を特定する経験的検証を続けるとともに、ヘイトスピーチと目される発話の言語的性質を概念的に明らかにする必要がある。具体的には、ヘイトスピーチと目される発話の（専ら発話者側の）性質のみならず、ヘイトスピーチの害の（発話者、聴衆、発話のターゲット集団、社会構成員等を横断して生じさせる）構成／発生機序を、他ならぬオンラインヘイトスピーチの特質に即して正確に把握する必要がある³⁴。

1.3 規制をめぐる困難② 解釈段階における文脈鋭敏性

1.3.1 難点3 規制対象を画する実践的難点としての語彙効果とそのフレーム依存性

さて他方で、（属性ではなく）表現の解釈上の特性に着目して規制に服すべきヘイトスピーチの限界を画することも実践的な難点を抱えている。詳細は次稿で検討するため、ここでは発話内・発話媒介的アプローチ双方を横断して生じる共通課題を、簡潔に指摘するに留める。

まず、ヘイトスピーチの外延をその発話内容のタイプから特定することは困難を伴う。「憎悪 hate」という発信側の事情に主として着目するか文字列上の見かけとは異なり、ヘイトスピーチは、表現内容を分析することで辞書的に（選言的に拡張したものであれ）定義される類の、特定の発話カテゴリーに属する発話集合として

は把握されないためである。

例えば、これを蔑称や隠語と比較してみよう。蔑称そのものは特定の記述の対象を指示するものではなく、さらには侮蔑的な発信者の意図を前提とするものでさえもない。よって、蔑称の発話による害は、蔑称の持つ意味内容を分析することから導き出されるのではなく、その受信者に対する社会的な効果（典型的には害を根拠とする禁止と禁止違反に対する諸制裁）から遡って把握される他はない³⁵。語彙そのものの効果は、その社会的な効果を支える文脈に依存する。ヘイトスピーチもこの点においては同様である。憎悪や害意なしにもヘイトスピーチ

は発話しうるが、それはヘイトスピーチの持つ意味内容からではなく、その受信者に対する社会的な効果から遡って把握されることを示唆する。すなわち、何が社会的に害を構成しあるいは発生させるかは、当該表現が置かれた歴史的背景・社会的状況を踏まえて把握されるだろう専ら現在の差別的慣行との関連性を有する広範な文脈に依存する。

この広範な文脈依存性から見えてくるのは、特定の発話内容を表現類型化し、地理的・文化的に多様である現在の差別的慣行と個々に結びつける作業の経験的な困難である。例えば、ある属性（例えば性別）に着目した蔑視的な表現が、当該集団に属する者一般を「格下げ subordinate」し、あるいは「構成する construct」効力を持つと主張されることがある。しかし、集団と個人との関係には、部分的なものから全面的なもの、着脱可能なものから不能なもの、可視的なものから不可視的なものまで様々な場合が想定され、一律に効力を想定することはできない。この効力を発話媒介行為として見た場合には、社会状況と個別の場面に依存したその実現過程の実証的裏付けに依存することとなり、その妥当性の検証なしには効力の有無・程度は判定不能となろう。あるいはこれを発話内行為として見た場合には、果たして発話者の権威や状況の支えによることなく効力を生じるのか、さらには一様にその効果を生み出すのかとの疑いが出されることとなる³⁶。

次に、発話のフレーム依存性が挙げられる。

1.3.2 小括

以上の小括とともに、上記難点が時間的な変

（受信者の訴えが認識上の端緒となりうることは当然としても）受信者に対する社会的な効果と言っても、個々の受信者にとって「そう聞こえた」ことや「そう読めた」ことからこの効果を直接に把握することはできない点が挙げられる。ましてや発話が特定の「メディア」を介してなされる場合には、上記文脈を形成するそのメディア特性を持つ効果も無視することはできない。言語や行動を成り立たせ、解釈させるフレームが重要となるのはここにおいてである³⁷。

もちろん、こうした先行フレームを誤解する、あるいはその複層性を特定フレームに引き寄せて誤用する事態は、物理メディア（銅像等の象徴的建造物等）³⁸、紙メディアにおいても起こってきたかもしれない³⁹。しかし、例えば漫画作品の特定のコマ中に現れる登場人物による蔑称や差別的言動は、（漫画の読み方に通じていない、あるいは当該作品が置かれた歴史的な文脈、当該ジャンルに通じていない読者にとっての語彙への直感的な反応を除いては）物語全体における当該登場人物の性質を表示する（その必然性から要請される）ものとして受容され、現実における蔑称としては一般的には機能しないし、そうした解釈が承認を受ける場面は（あるタイプの批評ではありうるが）限定的である⁴⁰。登場人物が語りかけている対象は、文字通りには現実のあなたではなく、その発話の害は物語において閉じるフレームの力学とともにあるが、こうした事態はオンライン空間における情報への接触の多くについても当てはまるだろう⁴¹。

容過程の中にも位置付けうることを確認する。

1.3.1で確認した通り、ヘイトスピーチは、当該発話がなされた歴史的背景・社会環境において、ターゲットとして名指された集団に属する個人に特殊な害をなす**機能的に把握された発話集合**である。ここでの害の候補としては、身体・生命への危惧感や名誉感情・社会的評判の毀損、民主的熟議の基盤の破壊、あるいは多義的であれ「尊厳」の毀損などが挙げられるが、1.2.2及び1.2.3の通り、その害の発生は（抽象的に把握することには困難を伴う）個別の関係性に依存した現実化段階での把握を要する性質を持ち、また1.2.4の通り、当該区別の改訂実践の中でのみその害が把握される種類のものであった。

こうした概念的な浮動性は、害を取り巻く社会レベル、個人レベルでの時間的な文脈変容にも影響を受ける。第一に、社会レベルで見た場合、上記のように機能的に把握される特性から、同じ表現内容であっても個別の社会環境によって発生する害の種別やその発生機序が異なるのみならず、当該社会環境自体の時間的推移に従って変動を被らざるをえない。第二に、個人レベルで見ても、特定集団の属性への帰属・コミットメントの長短、個々の生活史上の時間的位置により、個人のアイデンティティに結びつく態様もさまざまに異なる。

さらに1.1.2で確認したように、規制作用の限界は（法制定・解釈・適用の諸段階における）概括的な正統性付与をブロックする観点から画されるが、特定時点で作られた規制は（その時点を超えて）時間を通じて維持される固定性を発揮する。この結果、文脈上の変化に富んだ対象及び状況についての規制は過小・過剰な作用を發揮する傾向性を持つ。この過小・過剰作用を修正する権能を（一定の幅を持つ）解釈に委ねることもまた、当該解釈の不確実性を国家の特定機関（部門）が処理することで、別のリスクに転換する決定と等価なのである。

こうした時間を通じた害の発生・構成の諸段階を把握することなく、発話が構成・発生させる害を表現内容そのもの持つ害と取り違えることは、文脈に適合しない規制の誤作動を招くこととなる⁴²。以上より、ヘイトスピーチを（法）規制の観点から論じる際には、（当該ヘイトスピーチ自体の効果を決定する要素の正確な把握に加え、その維持・消失過程をも対象とする）時間を通じたコントロール手法を提示する必要がある⁴³。続く第2節では、オンラインヘイトスピーチ特有の害の機序に対応した、多元的な規制枠組み・非規制枠組みを検討する。

2、オンライン空間への法的介入とそのリスクトレードオフ

2.1. オンライン空間上の表現形式

2.1.1 コミュニケーション形式の特殊性

さて、SNSやオンライン・フォーラム等、オンライン上においても無論、前述の機能的な意味で把握されたヘイトスピーチに類した発話

はなされる。そのうちのいわゆる「特定」やつきまといなどリアル空間における行動や被害に繋がるいくつかのものは、従前から存在した（対

面状況で畏怖を与え、静穏な生活を害する⁴⁴類の)リアル空間における示威的なヘイトスピーチに近い性質を持つし、また著名人によるSNSでの継続的発信等いくつかのものは、雑誌やテレビといったマスメディア上における従前から存在した差別助長的なヘイトスピーチに近い性質を持つだろう。事実、(民事賠償はもとよりとして)明示的に特定個人に宛てられた前者に属するヘイトスピーチ群については、本邦でも脅迫罪・侮蔑罪等に当たる行為である以上、実際に刑事罰が科されてきている⁴⁵。オンラインを通じた個人に対する差別的・侮蔑的発言が、少なくとも人々に広く知られるに至る段階において特段の断りなく「ヘイトスピーチ」事例とされ、報道などがなされるのは、こうした事情によるものである。

しかし、オンライン上における差別的・侮蔑的発言の多くは、リアル空間の対面状況においてこれまで蓄積されてきた事例群⁴⁶とは性質を大きく異にする。それらは、(技術的には双方向的でありながらも)コミュニケーション上は相

2.1.2 メディア媒介的な空間特性

第二に、コミュニケーション形式のみならず、メディア媒介的なオンライン空間の空間特性にも着目できる。まず、オンライン空間の特性として、空間的な分割の困難性がある。これはオンライン空間においては全てが可視化されている(あるいは可視化されるべきである)が故にリアル空間を前提にした規制の拡張や規律の一様性を要求するという事態に直結するわけではなく、むしろ適切な空間的限定の下では正当であったはずの要求や法的介入が、その範囲・

互行為性を前提としないし、実際になされているやりとりの多くは相互行為の従来の定義を満たさない⁴⁷。この結果、オンラインヘイトスピーチの害もまた、対面状況におけるターゲット集団への攻撃に結びつく直接の害ではなく、プロパガンダ的ヘイトスピーチの流通・蔓延による(主には非ターゲット集団における)集団形成及び同集団内の差別・偏見の助長を通じた間接的な害のみを生じさせる⁴⁸。(現実の差別的集団がオンラインでも結集する場合や、特定の差別的フォーラム等で集団形成がなされる場合等の例外を除き)現実の集団維持・形成の契機を必ずしも伴わないオンライン上のヘイトスピーチの多くは、特定集団に紐づけられた危害の(相当の蓋然性のある)予期・把握を困難にする。他方、コミュニケーションの媒介者やアーキテクチャの介在とともに新たな害の発生・構成機序をも導く。次稿で詳述するものの、この相互のコミュニケーション形式の特殊性については、リアル空間・オンライン空間の議論を分かちものとして、第一に留意する必要がある。

権限を超えて全域に及びうという空間特性として把握される。

この典型事例として、鳥取県青少年条例による「指定」を介したネット書店からの書籍締め出し事例を比較対象に置くことができる。具体的には、危険な吹き矢・レーザー等の制作方法を内容に含む「アリエナイ工作シリーズ」等の書籍が青少年健全育成条例上の有害図書へ指定された後、結果としてオンライン書店 Amazon で販売ページからの取下げがなされ、鳥取県外

のユーザーに対しても事実上購入が困難となった事例である。(他でも検索可能な危険物制作方法を書籍に載せることと現実の危害との関係を掘り下げること、あるいは危険情報規制とポルノグラフィ規制との比較を行うことも有益であろうが) 本稿の関心からする主たる問題は、この下げがオンライン書店の自主的な判断と評価するには困難な事情がある点にこそある。すなわち、鳥取県条例が「インターネットの利用その他の方法により鳥取県内において」図書の販売を業とする者が有害図書販売行為を行った場合には罰金・懲役を課す旨を規定しており、オンライン書店としては、購入者の属性に応じて適用される(本邦だけで見ても最大2000弱の)地方自治体の条例の規定ごとに販売対象図書をカスタマイズする対応コストをかけるか、書籍自体をページから取り下げる(あるいはアダルト指定する)かの二択を迫られていたはずである。そもそも前者の実施にあたっては、オンライン書店側で購入者の正確な属性を確認・蓄積することが前提となる以上、事業の実施のために必要となる以上の情報収集をオンライン書店側に求める点で背理である。この点に鑑みれば、強いられた選択肢から選択している事実を以て、オンライン書店の自律的判断であるとみなすことには困難が伴うだろう⁴⁹。結果として、鳥取県の地理的範囲を超えて、書籍へのアクセス遮断は当該オンラ

2.1.3 「不快」の過剰とその抑止

第三に、表現が流通するSNSを念頭におけば、検索行為やフォロー行為を介して(原理的には可視化・開放されつつも)主として自らの意思の下において、さらにはアーキテクチャ上の

イン書店日本語サイト利用者全体に及ぶこととなった⁵⁰。まとめれば、オンライン空間においては従来考えられてきた分権性が機能不全に陥った事例として総括できる。

翻って、オンラインヘイトスピーチ規制についても、同様の地理的な拡張作用が付随する。第一に、対象の性質として、オンライン上のヘイトスピーチによる害が個別化されている場合、すなわち侮蔑・中傷や威圧、脅迫、プライバシーや名誉を脅かす虚偽の事実の流布、暴力の扇動等の行為群にあたる場合はリアル空間における害と機能上等しく、規制の拡張や遮断措置も可能になる⁵¹。一方、害が個別化されておらず、地理的(歴史的・文化的)条件と表現の具体的文脈が不確定の場合は、発話上、何らかの差別的な言辞を含んでいる発話がなされたとしても、(公人に対する批判や権力と結びつく「マジョリティ」属性への批判をも許されなくなる事態を想起すれば自ずと明らかのように)それだけでは規制されるべき対象の範囲確定はできない。第二に、規制の正統性確保の観点からは、規制作用の及ぶ地理的範囲が規制主体の地理的管轄に依存している点も指摘できよう⁵²。オンライン空間の分割困難性・全域性はリアル空間を前提にした規制の拡張や規律の一様性を導く条件ではなく、分権性の機能不全ゆえに極めて限定的な条件の下でしか規制を正当化しないのである。

制約において、いわば「閉じられている」特性も挙げられよう。

まず前者から確認する。オンライン空間の特性上、自らにとって「害と感じられるもの」に

接触する起点も（あるいは回避しなかったという事後対処においても）自らの意思によるところが相対的に大きい。発話の害が現実化するためにはそもそも（その表現内容以前に）発話が届ける身体との時間的・空間的な近接性が条件となるところ、オンライン空間に存在する発話についても、この時間的・空間的近接性の検討は不可欠となる。このプロセスを無視し、空間上に存在する言説一般を等しく投げかけられた「害」類似のものとして取扱うことは、不快原理（offence principle）が陥ると同じく過剰規制問題を生じさせることになる。

そもそも不快原理の提唱者であるジョエル・ファインバーグは同原理の適用にあたり、不快の中でも、程度として深刻かつ一般的なものに限定して問題となる不快を捉え、その大規模さや継続性に加えて、道理にかなった回避可能性（容易性）及びその不快さを同意の下で引き受けたか否かを考慮に入れていた⁵³。それと知りつつ自らの意思で近づいた不快、すなわち引き受けられた不快を危害類似のものとして構成するのは、危害に属する範囲を著しく拡大することになるためである⁵⁴。確かに検索やSNSを通じて「目に入る」という点においては非意思的な側面が強い（いわばテレビ的である）と言えなくもない一方で、文字・音声・動画を含めてコンテンツと向き合うプロセスにおいては回避や統御が可能な、意思的な側面が強い（いわば新聞・雑誌的である）⁵⁵。翻って非意思的な遭遇場面においても、続いて回避や意思的なブロックも通

常は可能であり、また支援されているだろう。

続いて、後者のアーキテクチャ上の制約、すなわちコミュニケーションを可能ならしめる媒介者・アーキテクチャを介した発信・受信プロセスについても確認する。まずオンライン空間上に存在する表現は、そもそも（発話の内容以前に）その発信段階においても受信段階においても複数の時点をもつ上に、（アクセス上の制約が一般的にはないという限度を超えて）必ずしも参加者全員が共有する「一つの」空間で共有されている表現ではないという特性をもつ⁵⁶。一見すると、トレンドを通じた関心のまとめあげやリアクション機能を通じたやり取り、あるいはいわゆる炎上事案の発生を通じた同時的な発言の集積は対面的同期性あるいは純粹同期性の拡張のように感得されるわけだが、（過去の事件や誤報をもとにした即時的炎上もしばしば生じる以上）この経験的側面はむしろ上記の擬似同期性を覆い隠すアーキテクチャを通じた選択の結果であり、対面状況とは隔絶した性質を持つ。次に、（トランザクションは別として）相互行為の前提となる相互認識は欠如しており、こちらが見ていることは同時的にはあちらには見られておらず、逆もまた然りである⁵⁷。すなわち、想像的には集合的でありながらも実践的には非相互的であり、（特定の「場」の規範を通じた相互調整という以上に）現実にもそうであると即時的・明示的には相手と共有することのない特性を有するのである⁵⁸。

2-2. オンライン空間での害の発生機序

2.2.1 害の多層性

以上を踏まえて、本稿の主題であるオンライン上のヘイトスピーチの害の多層的な現れを見る。第一に、上記 2.1.1. 及び 2.1.2. で確認したコミュニケーション形式の特殊性及び空間特性に由来する害の現れについてである。あなたがターゲットとなるマイノリティ集団の一人であると仮定しよう。リアル空間では自らの住まう概して平穏な生活上の世界に突如としてヘイトスピーチに属する発話群が畏怖を与える形で介入してくる⁵⁹。一方、オンライン空間における多くの発話群では、事情は異なる。まず、この介入は自らの意思が何らかの起点を成している社会的現実あるいは世界制作の結果として現出することとなる（害の意思的選択への依存性）。そして先に「メディア」媒介性と呼んだ効果の帰結として、対面状況では突如としておそってくる、畏怖を伴う（あるいは無用な議論に巻き込まれる）等々の性質は、非直接性及び物理的距離を介して失われる（害の間接性）。またそもそも発話とその受領とが、疑似同期的な錯覚を与えつつも、原理的には隔たっているという時間的距離を基礎として、各種遮断の経路・手段を複数選択しうることが多い（害の回避手段の選択可能性）⁶⁰。これを支える事情として、反対にリアル空間での署名（集団形成）やデモ、つきまといや粘着行為（あるいはこれらの回避）には多大なコストが付きものであったのに対して、オンライン空間での同様の行為への対抗措置もまた⁶¹、比較的に小さなコストで済む点も挙げられよう（害への対抗手段の選択可能性）。

以上のコミュニケーション形式の特殊性及び空間特性から、オンライン上におけるヘイトスピーチの多くは、「具体的に特定可能なターゲッ

トに対して、直接的な害を与えたり、あるいは個別的な害を与えたりするわけではない場合におけるヘイトスピーチ」⁶²に属する。少なくとも個別の発話単位でオンラインヘイトスピーチを見た場合、その「害」に基づく自由制約の正当化（危害原理）が機能不全に陥るのは、上記のコミュニケーション形式の特殊性及び空間特性の多元性・多層性を規制文脈に置くための「翻訳」に失敗してきたためである。

第二に、上記 2.1.3 で見た時間的特性に由来する害の現れについてである。オンライン空間に存在する一群の表現は、発信・受信の各の段階で複数の時間を含み持つ一方で、あたかもそれが同時的であるかのような外観を与える。ある彫像が「歴史的な文脈のなかに置けばむしろ人種解放と平等の象徴であるものを、独立の存在として瞬間的に見ることにより、まったく逆の意味のもとで理解」⁶³されうるように、当該表現がもたらす効果もまた同時的／異時的な複数の解釈を持つ。その瞬間にアクセス可能な表現に尽くされるわけではなく、時間を通じて様々な表現が積み上がるとともに、失われ、さらには忘却され、文脈を失い理解を阻むものも数多ある。この結果として、表現自体の脆弱性は、特定の表現が持つ、時点を通じた（異時的な）解釈のされ直しのプロセスそのものに及ぶことになる。たとえば、ヘイトスピーチを特定集団に属する対象者に認識させることがすなわち社会の中で平等な地位（尊厳）を持つ「安心」を奪うことだとする類の規制積極論⁶⁴と、法学的な言論の自由市場論を基礎としたヘイトスピーチ規制消極論とが少なからぬ緊張関係に立ったままで現在まできたのは、上記の理由に

基づくものである⁶⁵。

第三に、上記特性の上でオンラインヘイトスピーチを規制することに由来する害の現れについてである。上記の多層的なコミュニケーション構造・空間構造・時間的構造を持つオンライン空間では、(とりわけ万人に開かれた公共の場における同内容の発話とは質的に異なる以上) 発話による尊厳毀損の態様やその文脈もまた変容を被る。経験科学的研究が提供するエビデンスの匙加減⁶⁶で文脈依存性を持つ規制の濫用が容易にもたらされうる以上、上記 1.3.2 で見たように規則制定自体の恣意性の問題のみならず、規制を受ける特定の表現類型(及び個別の表現)への該当性判断及び(必ずしも国家に限られない)判断主体のに対する統制の問題を避けることは

2.2.2 オンライン上の「社会的暴虐」の過剰

もちろん、第三に挙げた表現に対する諸介入の目的及びその限界を論じてきた議論は、必ずしもオンライン空間に限られない。ただしその空間特性により、表現自体が持つ文脈の脆弱性と他者関係の中で表現の脆弱性が増幅され、その規制作用の限界を変動させる以上、ここで思想史上の議論として J. S. ミルの『自由論』を振り返ることは有益であろう。

かつて J. S. ミルは、『自由論』第 4 章において「他人に対する義務の問題」と「自己配慮に属する行為の諸問題」とを劃然と区別した⁷⁰。この区別の趣旨は、単に義務の発生原因を他者関係に限定し、あるいは第 1 章で提示されているように国家が個人・集団に強制的・統制的に介入する目的を「自己防衛」に限ること(危害原理及び危害除去のための公平分担ルール)のみで

できない。すなわち、議論が規制の限界に向けられている以上は、規制対象たる特定の表現類型(及び個別の表現)の持つ害は、その規制自体による害(及びそれらの除去)とともに論じられなくてはならない⁶⁷。このこともまた上記規制消極論を支えてきた背景として想起されよう⁶⁸。法による特定の害の除去と法による特有の害の生成とは、原理的に同時的であり、一方のリスクを除去することは他方のリスクを生み出すトレードオフ関係にあり、法による介入と不介入の間でもトレードオフ関係は生じる⁶⁹。これらの結果として、種々の害に関する二階の決断と、決断を統制する法特有の原理としてその(正当性のみならず)正統性が求められるのである。

はなく、より細部に及ぶ。仮にある行為や表現に何らかの害が存在するとしても、統治上の正統性をもって自らへの介入が許容される「他人に対する義務」を生じさせるのはいかなる害の種類であるのか、また「自己配慮に属する行為」といっても他者と接触しうる以上はいかなる介入(非介入)の統制が許容されるのか、がここでの主題である。ミルの見解では、前者については危害原理が妥当し、後者については更に中間に社会的な支えを要する「状況の多様性」⁷¹の保護が妥当する。つまり、専ら自らにのみ関わる愚かさ、墮落、低俗な好みといった性向、すなわちそれ自体としての価値に乏しく、ときに他者からの嫌悪を受けるだろう「醜い自由」⁷²の行使もまた(時にもっともな警告を善意から受け、あるいは低い評価を受けることを覚悟せ

ねばならないとしても) 制約されてはならない。「醜い自由」は専ら、精神的専制に由来する「魂の奴隷化」⁷³を排することにその根拠を持ち、状況の多様性の維持・促進の支えによってその存立の場を持つだろう⁷⁴。憐れみや嫌悪の対象であることを、怒りや怨恨の対象(すなわち「社会の敵」)であることと取り違えてはならない、とミルは言う。「自分の嫌悪する(他人の)行為を自分に対する権利の侵害」と取り違え、「自己の感情を傷つける暴行であるかのようにそれに憤る」⁷⁵ならば、その総合であるところの社会的非難は社会的な支えを失わせ、社会的暴虐の遂行となる⁷⁶。法規制を用いてであれ事実上の諸力(相対的に多数者としての圧力や諸資本を用いた権力であれ)を用いてであれ、上記の中間領域を締め出そうとする傾向性が働くことは、社会的専制の必然である。

ミルはこの事態への対抗として「差異が存在することそれ自体」の有益性を、それがもたらす結果の良し悪しやその予測とは無関係に尊重されるべきものとして主張していた。他者に対する説得や警告もまた常に可能である一方で、差異を生み出す個々人の個性の発露はそれら説得や警告が存立するための条件である以上、社会的な「状況の多様性」の支えを維持・促進し、少数者の声を響かせる目的に従い、それら説得

や警告の限界もまた画される。

この結果として、統治の正統な手段として法が用いられる際には、少なくとも国家の暴虐を防ぐための統制が(公開性や代議制といった安全策はもとよりとして)危害原理に服する形で求められる⁷⁷とともに、社会の暴虐から市民における多様な「差異が存在することそれ自体」を保護するための統制が「状況の多様性」保護の原理に資する形で求められるのである。言い換えれば、国家の統治作用を統御することで市民的基礎としての安全を保障する原理と市民的基礎を保護するために市民間のコンフリクト調停自体を統御する原理とは別の原理でありつつ、ともに自由の保障に向けられている。

このようにミルに従えば、仮に道徳的その他の意味において「醜い」表現であっても、法的な介入は無条件に妥当せず、その統制は上記の原理に従うことが導かれる。発信と受信とは、社会におけるアクセス可能な総量としては一方の増減がもう一方の増減には直結するため、その片方のみを享受することが構造的にできない。言論はこの対称性ゆえに、個人において培われる受忍の限度⁷⁸及び社会において培われる寛容の限界において絶えず挑戦を受けているのだ。

2.2.3. 小括 「キャンセル」とヘイトスピーチ規制との反転図式、あるいは「理念」としての地位

もちろん、上記の原理を特定の表現類型(さらには個別の表現)に適用する段階では、その運用において消極的な市民間の深刻なコンフリクトを容認する事態を招き、「状況の多様性」を掘り崩さないための措置もまた求められよう

(ヘイトスピーチ解消法における理念提示及びそれに基づく各種の政府言論もまたその例である⁷⁹)。ジェレミー・ベンサム曰く「悪政に対する安全策」は社会的・精神的専制にも及ぶ。ではその措置として、いかなるものが構想しうるか。これま

での議論を不適切・不品位発言へと「キャンセル」作用を及ぼすことについての議論及び言論の自由市場モデルの議論⁸⁰から再記述することで、第2節のまとめとしよう。

さて国家の権力にも社会的専制等にも抗するものとして、他者を害する「おそれ」のある表現であっても、それが「おそれ」にとどまる限りは、法制度の中に位置を占めるにあたり多様性の維持・促進という観点からの保護要請を欠くことができない。しかし、この観点は、保護と規制の間との緊張関係を生み出す。一方には、顕著な害が顕在化していない限りにおいて、相手の口を噤がせるための措置を要求するのではなく、むしろ更なる説得・批判によって対抗する「より多く」の多様な言葉を紡ぎ出す営みが推奨される、とする考えがある。他方、瀰漫的な害の「おそれ」は、ある属性の下でマジョリティではない側に集中しがちであり、多様な言論提示自体が弱化するとの懸念があり、こちらもまた相応の理はある。この緊張関係は、2.2.1の箇所で見たとおりである。

しかし表現の自由の保障が及ぶ範囲は、個別の表現が持つ危害の評価によってのみ限界が画されるわけではない。ある時点の、変わりゆく社会の、移ろいゆく文化の限界のうちにおいて「悪い」表現とそうでないものとを切り分け、前者の発言者を「キャンセル」し、あるいは表現そのものに対する「カードを切る」誘惑にかられてしまうまさにその場面においてこそ、自らの意見に論駁・反証しようとする表現の行使と表現を通じた公共的討論が保障されなければならない⁸¹。これがミルの主張である。上記切り分けや「キャンセル」に駆られる誘惑へと抵

抗し⁸²、翻って自らの判断が真理であるという想定を正当化するためには、意見の解釈を訂正する可能性に開かれていなければならないためである⁸³。

時に規制積極論として援用されるウォルドロンもまた、感情毀損と区別して尊厳毀損を取り出す作業の中で、実際にはこの問題の一部に答えている。ウォルドロンによれば、尊厳とは、共同体における平等なものとしての地位を指しており、その地位に見合う承認と扱いに対する要求を生み出す源泉である。ここでの地位とは理念であり、実体的な危害を構成する価値とは異なる。それ故、恥辱や侮蔑によって傷つけられた感覚・感情は、ヘイトスピーチの「害」たる尊厳の毀損の徴候に留まる⁸⁴。無論、そもそも相互的な尊重に等しく支えられることで成り立つ自尊心などの毀損は、社会的地位と結びつく重要なものであることは疑いない⁸⁵。しかし、仮に特定のアイデンティティ属性への攻撃を介して生じる不快の感覚・感情をより分けることなく、社会的な尊厳への毀損という事態と同一視するならば、「差異と不合意のただ中で社会が運営されることをはるかに困難に」⁸⁶する。必要なのは、特定のアイデンティティやそれを強いる共同体の拘束を脱しうる、少なくともそうである「かのような」自由を行使しうる社会的関係を絶えず新たに構築することであり、各々の間の差異と不合意を消去することではない⁸⁷。この結果、言論を発し、また言論を受ける地位が（事実上であれ）毀損された場合に、その格差を是正・回復する促進措置は法がなすべき任務に属する一方で、言論を発し、受ける地位のあらゆる差異の埋め合わせを禁止や規制

を通じて実現しようとすることは、法がなすべき任務には属さない。なぜなら、発信を阻害する要因を可視化・ブロックし、また促進する是正措置を競い合わせることは言論を発し受ける地位を対称にする一方で、禁止や規制を受けるべき属性やそれへの「節度を持った表現」⁸⁸という表現の主題・方法自体を主たる争点にすることは、言論を発し受ける地位をも非対称なものに固定するためである⁸⁹。

このように、「うち」なる脆弱な個としてのアイデンティティをそのまま政治の状況に交渉不能な危害として持ち込むことは、主権者の恣意とともに、「アイデンティティ政治の無責任さ」と集会的決定の挫折を招かざるをえない。ハードな効果を伴う⁹⁰ キャンセルカルチャー群が、特定の属性のみを切り出してそれを包含する同様の動きを攻撃とみなすに至るような場合⁹¹には、社会における議論継続のための中間領域、いわば個人の取り決めに最も重きを置く私的自治と、集団の取り決めに最も重きを置く民主主義との間に横たわる、取り決め困難な領域を破壊するに至る⁹²。その理由は端的に、そこに至った過程に対する責任の所在も、あるいはそこから将来に向けた公正な回復・和解の

途も、「キャンセル」という営為では明らかにすることがないためである。法による介入が、その内在的制約からごく例外にとどまるのはこの故である。受忍と寛容を涵養する社会への信頼がなければ、一見したところの消極的自由の保護は名目に終わり、容易に国家による積極的介入へと転化してしまう⁹³。再度ウォルドロン言葉を借りれば、「政治の状況」を直視しない「アイデンティティの政治の無責任さ」⁹⁴と集会的決定の挫折を、法は任務として加速させる役割を持たないのである。

このように、切り分けと「キャンセル」は、発言・発言者を結びつけるとともにその差異の発露を不可視化する⁹⁵ことで、発言者が社会の中で自らの誤りの「おそれ」に開かれ、自らの訂正を施しながら進む可能性を失わせる。単に欲望を地下化・増長させるのみならず、自他の差異のただ中で、自らを訂正することが可能な市民の尊厳ある一人として遇する努力を無化してしまうだろう⁹⁶。では、将来に向けた公正な回復・和解に向けた、促進的環境の維持・促進のための法的介入の限界とは何か。この点を、第3節で論じることしよう。

3、訂正と赦しの条件 対抗言論促進的環境

3.1 訂正可能な空間と促進的環境

マーサ・ヌスバウムは、その著書『感情と法』において「社会を人間の尊厳を守る場所にする役割を果たし、市民が恥辱とスティグマにさらされない生活を送れる」⁹⁷ 環境を、促進的環境と呼んでいた。ヌスバウムは、市民を等しい市

民として尊重するためには、多様な市民の存在と各々が信じる（人生についての理に適った）包括的教説を多元的に保障する環境が必要であるとした上で、貧困、差別・ヘイトクライム、性的マイノリティ、障害の諸例に挙げつつ、ヌス

バウムは「正常」という規格によって社会をデザインしたことで市民の一部を「脆弱」な者とのスティグマへと押し込める社会秩序は、秩序だった decent ものではなく、その是正のコストがバラバラな社会には要求されると主張した。必要なのは（社会的専制と同様に、完全性の専制⁹⁸を排し）「さまざまな障害を負った人が「巧妙な相互作用」の人生を生きることができる「行動を容易にする環境」を創造する」⁹⁹ことである。

この環境の指摘は（ヌสบaum自身も引用する通り）ミルの「状況の多様性」保障の議論を引き継いだ上で乗り越えようとしたものであり、「状況の多様性」を現実化するプロセスを含む解釈を施したものである。他方、ヌสบaumの関心はミルと異なり、市民間のコンフリクト調停が円滑に行われるために用いられる法的介入の統制原理を探究することには向けられていない。ここから、抵抗としての表現をエンパワーするとともに、議論継続を可能とする促進的環境形成のために、立法・行政・司法作用にどのように権限配分するか、（プロバイダ事業者やプラットフォーム等）私人を通じた法執行として許否する条件がなんであるかを明らかにすることが、法を用いてヘイトスピーチへの対応をなす際の中心課題となる。

3.2 環境を維持・促進する義務 シティズンシップの尊厳

時に J. S. ミルは言論の自由の核心を真理の発見に見たと要約される。しかしより正確には、その時々において通用する一般的見解とは異なる意見が一般的見解へと挑戦する議論継続（訂正過程）による真理想定の正当化の内に見

確かに現在、表現の自由の武器化や歴史問題の司法化¹⁰⁰の功罪も盛んに論じられている。「正しさ」を決着する場が裁判であるの「臆見」（やそうしたアピール）が通用する場面において、一見したところの表現の自由の「正当」な行使に、様々な暴力性への加担を見てとることは（それら議論が法学上妥当か否かは置いておくが）不可能ではないかもしれない。しかし、歴史問題の司法化批判が、そもそも司法における「武器」となることへの懸念やその除去要求ではなく、そもそも歴史形成の場の中心は司法ではなく、歴史形成に不断に関与するのは社会の役割であると強調していたことが、まさにここで想起しうる。

上記 2.2.3 で確認した通り、表現の自由が十全に保障された空間を求めることとは、自らの表現が聞き取られる価値があるものとして受け止められる空間を求めることでも、他者の感情が毀損されうる「おそれ」を欠いた安全な空間を求めることではない。表現の対称性に鑑みればむしろ反対に、自らを傷つけるかもしれない相手の表現に向き合う環境を社会として支えらるとともに、「より多くの」表現が響き渡る空間を危害の除去のための負担として担うことで、自らの意見及びその解釈を訂正する可能性に開かれる必要がある。

た、という方が正確である。これは二つの解釈を持つ。第一には文字通りに、その時に正しいものとして通用している意見が誤っている可能性に開かれ（真理発見的側面）、意見の正しさを反対論に対抗することで明瞭に理解する（真理

確証的側面)のみならず、異なる意見がそれぞれに正しさを分有している場合にその両者を保存する(真理保存的側面)側面に、その核心たる所以を持つというものである¹⁰¹。しかしミルが同じ程度に重きを置いていたのは、そうした真理をめぐる発見・確証・保存の営為が特定の科学者や思想家の占有物ではなく、「普通の人間を彼らの到達できる限りの精神的高度に到達することを得せしめる」¹⁰²のために必須の要件であり、それこそが言論の自由の核心たる所以を持つという考えであった。

前者は社会功利に通じる真理の側面として主に『自由論』第2章で論じられる「真理に基づく正当化」に、後者の自己開発的側面は主に『自由論』第4章で論じられる「人格に基づく正当化」に区分されることがある¹⁰³。しかし、上記引用箇所近接性が示すように、議論を立て、また引き受けことは、この両方を横断している。自身のイニシアティブで意見が提起しえ、議論が起こしうるといふ信頼は、挑戦された一般の見解を擁護する者に合理的根拠とともに理解する契機を与えるとともに、強いられた教説としてではなく市民各人の人格の発露を通じた確信を成長させる助けとなる。したがって、ミルによれば、社会に対する真実を巡る挑戦はまずもって(ミルの想定と異なり、挑戦を受けた当人が矢面に立つか否かについては本稿の関心上は差し当たりは別の可能性を留保するとしても)議論に参加することで受け止められ、結果としてその度ごとに真実は訂正と確証に開かれねばならないことに帰結しよう。

無論、社会における平等な地位へと疑問符が付され、あるターゲット集団に対して自尊心を

蝕む「おそれ」が蔓延したときには、ターゲット集団からの議論提起・議論継続(訂正過程)は蓋然的に見込み難い。しかし、ミルの上記箇所によれば、社会の防御として採りうる方法は、ある種の表現をスケープゴートに捧げ、事前に刈り取る営為ではなく、むしろ「おそれ」のある表現の受け手をエンパワメントし、「安心」を再構築する努力という形式をとる。発話による害の「おそれ」が、単なる不快を超えて実際の害となるか否かは、まさしくそうした「おそれ」を取り巻く市民相互でその「おそれ」を受け止めうる支えに依存しているためである。

ウォルドロンは『ヘイトスピーチという危害』を著した翌年の2013年、別の論文において、人間の尊厳とシティズンシップの尊厳を並行的に論じていた¹⁰⁴。シティズンシップは、それぞれの政治体における、特別でありかつ平等で広範な社会的地位という、一見すると相反する要素を含む。シティズンシップは歴史を通じて変化し、場所によっても異なるが、いかなるシティズンシップもひとかたまりの権利・権力・義務・責任から成り立ち、その内容や適用は法的問題として定められる点では共通する。シティズンシップの尊厳は、特権階級の特別な権利ではなく万人への普及が目指されるが、それは同時に自らの利益を超え、全員の利益に公正に応える政策・法律・枠組みに注意を向ける義務を引き受ける人に向けた尊敬を含む。すなわちシティズンシップが尊厳たるのは、他者の尊厳を尊重する義務、法に服従し政治的敗北を受け入れる義務、市民権を能動的に行使する責任に結びつくためである¹⁰⁵。

責任と結びつくそのエンパワメントは、単に

マイノリティ成員を脆弱的存在として遇することによって社会から危険をあらかじめ除去し、あるいは成員を危険を拭えないという理由で社会的営為から「隔離」し、保護されるべき存在に留めおくこととは異なる¹⁰⁶。ある属性に基づくマイノリティ成員は（他者からのスティグマ

の帰結であれ、「自己スティグマ」¹⁰⁷の帰結であれ）一方的に保護されるべき「脆弱」な主体ではなく、むしろそれを「脆弱」化させてきた正常性・完全性に基づく社会秩序¹⁰⁸を変更することによって、相互協力し合う社会のメンバーとなることが期待されるだろう。

3.3 おわりに リスクのキャンセルから、リスク含みの選択のエンパワメントへ

「隔離」から自律的な離脱／再参入の保障へとシフトしつつある法の力点移動を踏まえるならば、「おそれ」に満ちた中でもコミュニケーションに参入するための「信頼」を構築することに、法の営為もまた向けられる。逆説的なことに、尊厳に満ちた社会を動揺させる「おそれ」から類型的にオンラインヘイトスピーチを刑罰をもって処罰するとは、シティズンシップの尊厳を普及することを止め、飛び地としての「二級市民」を許容することと表裏一体である。

確かに被害は、社会に埋め込まれた特定のアイデンティティ構成的属性への毀損を介し、局所化されて生じうる。まさにその場面では、声を上げる人を具体的に一時的には保護し、承認を回復することが求められるかもしれない。しかし、保護・承認に向けて論争を回避することを含む「キャンセル」作用は正統性を伴った法規制には例外的にしか含まれない。市民一人ひとりの脆弱さと相互依存の上でなお、潜在的には常に生じうる論争的な表現とともに立ち向かう、連帯とエンパワメントによって、シティズンシップの尊厳を各アクターが発揮する環境形成に、法のコントロールは向けられている。少なくとも現実の属性を直接に反映することがないオンライン上の表現空間においては、アイデ

ンティティ構成的属性へと向けられたオンラインヘイトスピーチとその害を抽象化したまま規制の議論を行うことなく、複数のアイデンティティを構築しつつある個別の人格を、尊厳と自尊心に満ちた、複数的な存在へと変じるプロセスへと開くことが求められる。

以上、本稿では、「シティズンシップの尊厳」を潜在的にもつものとは、武装解除を求めるのではなく、武器にも耐えられる「タフ」な空間を求め、自他の訂正可能性に開かれつつそれを絶えず確証し続ける存在である、という展望を見た。ただし、オンラインヘイトスピーチをめぐる状況には害をめぐる「エビデンス」の過剰利用の危険が常につきまとい、それが感情的なバックラッシュの一因となることも想定しやすい。こうしたオンライン空間における尊厳促進的環境としての法と、規制による、また規制をめぐる感情のわだかまりを解きほぐすのは困難を極めるが、この調停はいかにして可能か。また上記の展望自体が、将来における断罪の日が来ることを受け入れつつも、将来における赦しを期待する、寛容に満ちた社会の像の共有に資するだろうか¹⁰⁹。各種リスクトレードオフのただ中で「法を用いてものごとをなす」法哲学的に見たオンラインヘイトスピーチ規制の指導

理念は、シティズンシップの尊厳を育むだろう といを次稿の導きとして、一旦筆をおく。

「信頼の精神」¹¹⁰ であり続けられるだろうか。

以上

本稿自体への将来の訂正を期しつつ、上記の問

研究助成

本研究は JSPS 科研費 23K12354 の助成を受けたものです。

謝辞

本稿ドラフト版を基礎に、東京大学大学院学際情報学府における講義「学際情報学概論Ⅰ」(2023)、「学際情報学概論Ⅱ」(2022)等の講義を行ってきた。受講生諸氏の意見は、いずれも本稿校了への精神的な後押しとなった。個別のフィードバックの日を期しつつ、まずはあつく感謝を申し上げる。

註

¹ ヘイトスピーチという言語現象に対してその言語内行為性から分析した代表的な論考には、Rae Langton “Speech Acts and Unspeakable Acts,” (1993) やそれを引き継ぐ Maitra and McGowan “Subordinating Speech” (2012) がある。これらを筆頭に、現在に至るまで主としてフェミニズム哲学の背景の下で各種の分析が進められてきた。

とりわけ近時、Mary Kate McGowan による文字通りに両義的な名前を持つ著作 *Just Words: On Speech and Hidden Harm* (2019) を通じて広範な領域から論じられるようになっていく。直近では、特に同書をめぐっての論文集 *Australasian Philosophical Review*, 5(2) (2022) の Sarah Sorial & Sam Shpall, Mary Kate McGowan, Ishani Maitra, Robert Mark Simpson, Jeremy Waldron, Caroline West らによる各論稿及びそれらへのリプライは、次稿においても第2節を中心として明示的に参照することとなる。加えて、Daniel Fogal らが編じた *New Work on Speech Acts* (2018) の各論考も有益な示唆に富む。

² 本稿は、2022年3月の科研費報告会「オンライン（特にSMP）上のヘイトスピーチ対策と法的介入のグラデーション」（司会：本多康作）を修正の上で論文化したものである。同科研費報告会は、前掲注1）の論文集にも寄稿している Robert Mark Simpson をこれに先だつ2年前の2020年2月に本邦に招き、まさに “Is Hate Speech Punishable?” と題して開催した科研費研究会（2020, 司会：永石尚也）に由来した論文集構想報告会を兼ねたものである（この時の Simpson 報告は、近刊『差別・ヘイトスピーチ・法』（2023 発刊予定）第二部に邦訳「ヘイトスピーチは罰すべきか、罰に値するか」（永石尚也・本多康作 訳）として収録予定である）。

同タイトルに表れている通り、法領域の中でも刑罰に着眼したものであったことを引き継いでいる。例えば公の施設の利用可否をめぐる諸課題や、SNS 型のデジタルプラットフォームによるコンテンツモデレーションやアカウント削除等をめぐる諸課題等、多くの論点は別稿に譲らざるを得ない。この限定はもちろん国家による権限行使における刑罰という制裁の特殊性を前提としているが、2022年2月に最高裁が示した事例とも関連する。「大阪市ヘイトスピーチへの対処に関する条例」(2016年)に基づく公金支出（審査会委員報酬支出）の合憲性が問われた住民訴訟（最判令和4年2月15日）では、その合憲性の理由の主たる部分として、表現を制約する手段としては拡散防止措置としての掲示物等の撤去・オンラインコンテンツの削除要請及び認識等公表など、刑罰に至らない手段ゆえに必要最小限度の制約であるとして憲法21条1項に適合するものとされた。更なる比較対象として、本邦でも唯一例外的に罰則（12条）を規定する「川崎市差別のないまち人権条例」において、罰則対象をあくまでもリアル空間における程度の甚だしい行為へと絞り込むとともに、手続上も違反者への勧告、勧告に従わない場合の命令、命令に違反した場合の公表と罰則（罰金）という一連のプロセスを整備している点も参考にならう。

³ 予告的に、オンラインヘイトスピーチの害をその言語行為上の特性から分析することを旨とする次稿での議論を、本稿と関連する限りで記す。冒頭では、ヘイトスピーチによる害を因果的害と構成的害とに区分した上で、その発生機序を発語媒介のアプローチと発語内的アプローチのそれぞれから分析することを示す。第1節では、ヘイトスピーチがもつ発語媒介効果に着目した分析を、オンライン／非オンライン状況の比較において施す。この結果として、とりわけオンライン上においては、発語文脈の多様性と規制主体の多層化、そして発話者と宛先となる対象者との時間的・空間的・関係的「距離」により、規制の基礎となる明確な害悪は疑わしいことを導く。次いで第2節では、発話内行為としてのヘイトスピーチが持つ効果に着目した McGowan

による分析を、そのアイデアの元となった David Lewis の scorekeeping 論とその現代的解釈を与えた Derek Ball の議論によって修正を施しつつ、オンライン状況へと適用する。ヘイトスピーチがオンライン上においてなされる場合、時間的・空間的・関係的な「距離」により、会話において許された状況的規範の定立・適用は、多くの場合十全に作動しない。むしろオンライン上においては、規範定立の不発を前提に、当該発話への非難や通報など対抗的な議論喚起もまた惹起することで、当該社会における差別的視線と共に「差別への（対抗的）視線」をも可視化することになることが示唆される。以上の発語媒介的アプローチと発語内的アプローチに基づく分析を踏まえ、第3節では、オンラインヘイトスピーチに属する発話がなされた場面においてそれへの対抗言論を行使する際、（解釈上その他）諸資源の不均衡に鑑みた格差の害という3つの害について明らかにする。これは、直接的な規制を支持するというよりも、共生に向けられた環境形成の指導理念としての社会的な「尊厳」の概念（平等な地位、自尊心、シティズンシップ）を支える法の役割を示すこととなる。

- 4 京都地裁令和4年8月30日（非現住建造物放火罪）は、行為及び結果としては通常の犯罪と同一であったとしても差別的な動機が動機の発現が量刑上も考慮された事例として紹介される。判例解説として、金尚均（2022）「一連の犯罪について差別的動機が量刑上考慮された事例」を参照せよ。
- 5 辻田真佐憲・西田亮介（2021）『新プロバガンダ論』（株式会社ゲンロン）。なお両氏も想定しているだろうように、「耳を傾けすぎる政府」問題は決して日本特有の事態ではなく、SNSによって政府への賛否を含む意見が可視化される場所では見られる現象である。例えば権威主義的立憲主義と目されるシンガポールにおいても同様の事情は観察される。この点につき永石尚也（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」第1節を参照せよ。
- 6 曾我部真裕（2020）「インターネット上の情報流通と民主政」法学セミナー No.788などを参照せよ。本邦では広告規制などを介した不均衡の是正は議論されてきたが、表立って偽情報対策が現行法上の規律には乗っていない。これは当然ながら、政治過程そのものが一見非科学的に見えるものであれ特定の利益や信条の発露を真つ向から否定するものではなく、それを前提にしつつ正当性・正統性の調達の回路を複層的に構築する営みであるためである。無論、量的な限度や誤導的な情報の害悪がある以上は、その介入の限界についての問題提起は種々になされている。
- 7 ハーマン・カペレン & ジョシュ・ディーバー（2022）『バッド・ランゲージ：悪い言葉の哲学入門』（勁草書房・葛谷 潤、杉本英太、仲宗根 勝仁、中根 杏樹、藤川 直也 訳）第3節 3-2-2、Faulkner(2007)から引かれている箇所での例を見ることができる。ナチスの追手からある者を匿っている家主が、誰かを匿っていないか、と尋ねられた時に「いいえ」と嘘を述べることは、結果としてはその者に利する蓋然性が高いだろう。さらに真実を述べつつ嘘をつきうることを示す例としては、上記の例で追手に家主が尋ねられる直前に匿っていた者が窓から逃げ出していた、という例を提示できる。さらに、自らが真ではないと信じる事実を他者に真だと信じさせる点で、嘘と欺きは共通点を持つが、いずれもその後の事態の推移と相手からの信頼への違背継続などに対する規範的評価という事後の事実によって、その「害」が把握されることになろう。最後に、フィクション作品において内容上妥当する「フィクションにおける真実」（デイヴィッド・ルイス）は、その真偽判断において参照されるべき世界が現実世界と異なる特定の虚構世界にあり、その信念を抱くことは単なる現実世界との不一致という意味での偽の信念を抱くこととは異なる点もここで指摘できる。こうした「フィクションにおける真実」の多層性は、自身とは異なる他の人生や final vocabulary に対する尊重や寛容を育む点においても有益なものでありうるものと思われる（リチャード・ローティ（2000）『偶然性・アイロニー・連帯』斎藤純一・山岡龍一・大川正彦訳（原著1989年）153頁）。ウォルドロン の議論に即し、次稿でも詳述する。
- 8 大阪高裁令和3年11月18日はリアル空間の職場でなされた文書配布について賠償請求が認められた事例であるが、「●●人は死ぬ」といった文言やその思考を「野生動物」に喩える表現をヘイトスピーチと認定した一方で、「売国奴」などの表現はヘイトスピーチには当たらないが差別を扇動し、分断を強化する侮蔑的発話であるとしている。その上で、こうした文書の配布あるいはその放置がともに職場環境配慮義務に違反するものとし、その同時発生性を認めている。武井寛（2022）「使用者による差別的言作文書配布（ヘイトスピーチ）と職場環境配慮義務 /（大阪高裁令和3年11月18日）」を参照せよ。なお本事案において「労働条件」に職場環境配慮を含まなかった点を問題視するものとして橋本陽子（2022）「使用者による国籍差別的言動（ヘイトスピーチ）の違法性——フジ住宅事件」がある。
- 9 同様の区分として、Matthew Kramer (2021), Freedom of Expression as Self-Restraint 第2章 [Components of a Theory of Freedom of Expression] を参照せよ。ここでは統治上課される自制として、4つの中立性、すなわちコミュニケーションの形式、内容（主題・観点）、主体によって取扱いを変えないという中立性と、行為との結びつきが偶然的である（行為からの独立性をもつ）コミュニケーションに対して干渉を控えるという中立性とが要求されるとしている。仮にそれらの害が複合的に生じる蓋然性が高いと考え、その複合性を規制根拠とする場合には、まさにその分ち難さを生ずる場面を厳格に特定し、そうでない場面

から分離する必要があるためである。

- ¹⁰ やや異なる文脈ではあるものの、横大道聡（2013）『現代国家における表現の自由—言論市場への国家の積極的関与とその憲法的統制』第8章を参照せよ。特に、事業者の自主的取り組みを促すとともにその過剰さによる不利益を回避させる統制された介入が、ここで着目に値する。フレデリック・シャウアーは論文“Principles, Institutions, and the First Amendment”（1998）において、芸術、図書館、大学、公共放送ネットワーク等の様々な文化制度について、制度それぞれの規範・法理を検討する必要を論じ、制度特定の修正1条論の構築を提唱していた。こうした文化制度が本来的に果たす機能の侵害を（部分社会の再来の危険を持つ「専門家の自律」保障ではなく）「公共討論への違憲な介入」という枠組みから捉えることを、横大道は主張する。なお、そこでは「あるべき公共討論」が措定されるのではなく、制度の本来的役割への阻害が「あるべきではない公共討論」への転化として捉えられている点にも、今後の検討における本稿との接続を期待することができる。
- ¹¹ 総務省（2022年8月）「プラットフォームサービスに関する研究会 第二次とりまとめ」における違法・有害情報対策とともに、利用者情報の取り扱いを参照せよ。権利侵害情報についてはプロバイダ責任制限法及び関係ガイドラインにより、その他違法な情報についてはインターネット上の違法な情報への対応に対するガイドラインにより、違法ではないが有害な情報については違法・有害情報への対応等に関する契約約款モデル条項により、青少年に有害な情報については青少年インターネット環境整備法によるフィルタリング等によって対応するものと整理されている。発信者情報開示に関して、総務省「発信者情報開示の在り方に関する研究会最終取りまとめ」（2020年12月）を参照せよ。
- ¹² 「サンタクララ原則—コンテンツモデレーションにおける透明性と説明責任について」（2018策定、2021改訂）でも、国家が事業者によってモデレーション判断に関与することは排除される一方で、事業者による恣意的な判断を回避すべくアカウントビリティ確保のための原則が列挙されている。あわせてアメリカにおける通信品位法230条(c)(2)(A)の免責のための要件をめぐるオンラインの検閲の防止に係る大統領令（2020年5月28日）及び司法省勧告（2020年6月17日）を参照せよ。同免責の射程を限定する各種否認法理については、平野晋（2014）「免責否認の法理（『通信品位法』230条）」総務省・情報通信政策研究所を参照せよ。
- ¹³ 本邦において特徴的な例としては、川崎市における「川崎市差別のない人権尊重のまちづくり条例」における、インターネット上のコンテンツに対するインターネット拡散防止措置（プロバイダ等への削除要請）がある。法令上の効力として削除要請に応じない場合の制裁は伴わない点では、その効果としてはヘイトスピーチのない社会の実現に向けた市の認識を提示する象徴的效果や表現内容等の情報提供機能に留まり、また手続的に学識経験者により構成される審査会の意見を踏まえるものとして評価できる。他方でこうした「要請」にこそ、潜在的な個々の表現者にとっての萎縮効果を排した明確性とアカウントビリティが求められる点は、本文記載のとおりである。法における正統性確保を支える「誠首」とは異なる正統化原理の動揺については、興津征雄「正統性の構造分析—行政国家の正統性を手がかりに（上）」法律時報93巻1号及び同論考「（下）」法律時報93巻2号を参照せよ。
- ¹⁴ 前掲注9）Matthew Kramer（2021）第2章 [Components of a Theory of Freedom of Expression]
- ¹⁵ 最判令和4年2月15日の「大阪市ヘイトスピーチへの対処に関する条例」（2016年）に基づく公金支出（審査会委員報酬支出）の合憲性が問われた住民訴訟の事例では、人種差別的な意識の助長をもたらす言動についての抑止の必要性から、その詳細に踏み入ることなく規定の合憲性を導いている。しかしこれが法的強制力を伴う刑罰を付加するケースにおいて妥当するかは疑わしく、にもかかわらず法的介入を要求するヘイトスピーチについての限定を明示的には施さなかった点については不足がある。この点に関連して、本邦における裁判所の審査手法をヘイトスピーチ規制に対して詳細に検討したものとして、Keigo Obayashi（2021）“The free speech jurisprudence in Japan: the influence of comparative constitutional law,” in (ed.) Shinji Higaki, Yuji Nasu, Hate speech in Japan : the possibility of a non-regulatory approach に詳しい。
- ¹⁶ いわゆる国際人権規約B規約（1966年採択）及び人種差別撤廃条約（1965年採択）には「ヘイトスピーチ」という語こそ表れないものの、差別的取扱いの禁止及び同取扱いを助長する表現についての禁止を、その制定時より盛り込んできた。例えば、国際人権規約B規約第2条第1項は「この規約の各締約国は、その領域内にあり、かつ、その管轄の下にあるすべての個人に対し、人種、皮膚の色、性、言語、宗教、政治的意見その他の意見、国民的若しくは社会的出身、財産、出生又は他の地位等によるいかなる差別もなしにこの規約において認められる権利を尊重し及び確保することを約束する」とし、表現の自由の一連の規定中における第20条第2項において「差別、敵意又は暴力の扇動となる国民的、人種的又は宗教的憎悪の唱道は、法律で禁止する」と定められる。後者の人種差別撤廃条約においては、その第4条において「締約国は、一人種の優越性若しくは一の皮膚の色若しくは種族的出身の人の集団の優越性の思想若しくは理論に基づくあらゆる宣伝及び団体又は人種的憎悪及び人種差別（形態のいかなるを問わない。）を正当化し若しくは助長することを企てるあらゆる宣伝及び団体を非難し、また、こ

のような差別のあらゆる扇動又は行為を根絶することを目的とする迅速かつ積極的な措置をとることを約束する」との規定がなされている。その上で、国連人種差別撤廃委員会が、ヘイトスピーチに関する一般的勧告 No.35 (2013) において、人種差別撤廃条約第 4 条が民族的・宗教的属性に基づくヘイトスピーチの禁止を含む趣旨であることについて明示的に言及している。文言を含む歴史的な経緯については、木下昌彦 (2020) 「Hate Speech in Japan を読む 国際法・憲法・比較憲法の観点から」に詳しい。

¹⁷ 国連人種差別撤廃委員会のヘイトスピーチに関する一般的勧告 No.35 でも、犯罪として処罰することについては人種主義等の最も深刻な発言に留め、それに至らない発言については他の手段をとるように勧告されている。

¹⁸ Sarah Sorial & Sam Shpall(2022)”” 第 6 節では同法の運用実態に照らせば、規律が苛烈なものとなる場面は決して多いものではなく、主として法の表出効果に期待しているものであるだろうことを指摘している。しかし、こうした社会的状況やオンライン空間の技術的環境の変化を反映させない規律を下に、個々の表現の当否の判断を専らその時々の人権委員会の判断という手続過程や、オンライン空間であれば私的主体でありかつ媒介者であるところのプラットフォームを通じたコンテンツモデレーション、すなわち事実上の権力へと委ねるおそれを認識する必要はあるだろう。この点につき、永石尚也 (2019) 「法の執行と正統性」『法政策の試み第 19 集』(信山社) の第 2 節を参照せよ。

¹⁹ Ayako Hatano(2021) “Hate Speech and International Law: The Internalisation of International Human Rights in Japan,” in (ed.) Shinji Higaki, Yuji Nasu. Hate speech in Japan : the possibility of a non-regulatory approach では、本文に述べたように本法のヘイトスピーチ解消法が文言上の刑罰の抑制性や非包摂性において人種差別撤廃条約の国内法化としては不十分である一方で、立法過程での広範なステークホルダーの関与や社会における議論喚起を通じて、立法後のヘイトデモ件数の急激な減少との相関を持つ点を指摘している。

²⁰ 前掲注 7) 『バッド・ランゲージ』第 5 章 [概念工学] を参照せよ。この上で、意味変化のコントロール可能性が問題となろうが、本稿では意味変化要因とコントロールとがマクロ的な歴史的・社会的状況に相関的であると考えており、この点につき意味変化要因の把握不能性とコントロール可能性の欠如を訴えるカペレンの「儉約フレームワーク」の説明 (同書 5-3-3-2) と本稿の立場とは軌を一にする。

²¹ 本文で述べた事情に加え、さらに当該属性と現実集団との間にある歴史的過程にも着目できる。過去から継承された様々な属性に付随した抑圧があることは事実としてある。他方、それら抑圧は各々が質的に異なるとともに、それらを支えた当時の経緯やイデオロギーもまた統一性を欠き、さらには過去のある時点から現在に至る回復や和解、その後の「赦し」の歴史的過程及びその実践も様々であることに留意する必要がある。林志弦 (2022) 『犠牲者意識ナショナリズム』(東洋経済新報社・澤田克己 訳) の言葉を借りれば、グローバルな「記憶の戦争」の問題が表面化しているとも言えるだろう。ホロコーストが人類史的な災禍として認識されている事実には疑いはないが、一方では「ホロコーストのコスモポリタン化」現象と呼ばれる脱地理化のねじれがあり、他方では他の民族的・宗教的虐殺がそうみなされているわけではない (同時代の「西側」を含む各国において、身体障害・精神障害を持つ者に対する虐殺や隔離がこれと並行していたことを思い起こそう) 緊張関係を抱えている。このように何が人類史的な災禍と見做されるかをめぐる林日くの「記憶のポリティクス」がここにあらわれる。

また国際人権規約 B 規約や差別撤廃条約の文言修正の経緯において、ヘイトスピーチに対する刑罰化を強く主張したのが専ら「東側」諸国であり、「西側」諸国が (憎悪言動の唱導の排除の重要性に鑑みても) 制裁を背景とする禁止がもたらす負の効果の観点から例外を設けさせた経緯を、ここで思い起こすことができる。この経緯は、各人権類型の間の緊張関係を内包する統合的な人権保障の文脈においても、「記憶のポリティクス」の動きとそれが規定になった後に普遍的なものとして通用してしまう事実を取り込む必要性が明らかであることを示す範例となる。この経緯につき、前掲注 16) 木下昌彦 (2020) 91 頁以降を参照せよ。

²² Judith Butler に代表されるこうした本文記載の見解に対し、サリー・ハスランガーは規範的含意を受け入れた上で「抑圧の構造を強化するのではなく、むしろ切り崩す助けになることが見込まれている」ものとして、誤ったカテゴリーに対しても改良的なアプローチをとる優位性を主張していた。しかし、この「切り崩し」の有効性については、前掲注 20) に記載した「儉約フレームワーク」との関連で筆者としてはやや懐疑的である。サリー・ハスランガー (2022) 「ジェンダーと人種——ジェンダーと人種とは何か? 私たちはそれらが何であってほしいのか?」(慶應義塾大学出版会、木下頌子訳) を参照せよ。なお司法上の判断としてはそれゆえに、例えば平等条項の適用において、当該属性に基づいて別異の取扱いをする傾向性を緩和・除去するために、当該属性をブラインドし、当該属性を理由とする取扱い上の不平等を是正することを目指すというように、当該誤ったカテゴリーについては専ら消極的に、定義不能な属性として遇してきたとも捉えうる。

²³ 荒井ひろみ, 和泉悠, 朱喜哲, 仲宗根勝仁, 谷中瞳 (2021) 「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」言語処理学会 第27回年次大会 発表論文集などを参照せよ。このように実際、コンテンツモデレーションの実施や判定基準の正当性確保に向けた調査として、多数の発話タイプを用意した上で、現にその言葉を見たときに直感的に「ヘイトスピーチ」に当たると考えるか否かを多数の被験者に判定させる調査がある。より具体的には、地域「独自の歴史・地域性を踏まえたヘイトスピーチ理解と、人文社会科学におけるヘイトスピーチに関する理論的知見に依拠したアノテーションガイドラインを設計」した上で、結果の統計的処理によって、当該地域、当該文化、当該時点、当該発信/受信メディアにおいて広く認識されるだろうヘイトスピーチ該当性とその濃淡を付与する、というものである。当該論文でも、語彙そのものが特定の文化カテゴリーと結びついている場合の偏見の混入などの限界が指摘されるところであるが、より深刻なところとしてはアノテーターの属性を問わずにこの調査を行う点ではなかろうか。この場合には、特定マイノリティへの感受性や共感性の有無によるバイアスが混入するため、ロナルド・ドゥオーキンが『権利論』で提示した「外的嗜好」と同様の問題が生じることが懸念される。

- ²⁴ 安藤馨・大屋雄裕 (2017) 『法哲学と法哲学の対話』(有斐閣) 第5章における「応答」末尾の問題提起を参照せよ。なお、ジェレミー・ウォルドロンがヘイトスピーチ規制における属性の無制限の拡張を問題視し、「アイデンティティ政治の無責任さ」と集会的決定の挫折と評していたことが想起される。
- ²⁵ 例えば沖縄県における「沖縄県差別のない人権尊重社会づくり条例」(2023)においては「県民」という属性に着目して、その地域社会からの排除煽動などが生じた場合における言動解消措置を講じることを定めている。しかし、出自でさえない「県民」性の範囲が問題になるとともに、(もちろんその歴史性に鑑み、他の県と異なるカテゴリーの特殊性は認めつつも)「県民」と制度的な地方自治体という公的・政治的アクター性と密接不可分の結びつきによって、有用な表現・議論をも縮小させかねない懸念についていかなる対処がなされているのかは、以後3年ごとの見直し運用において必然的に争点となるだろう。
- ²⁶ こうした点は、ときに原理的な「証言者の不在」(ジャック・デリダ)あるいは「ホロコーストの表象不可能性」(ヘイドン・ホワイト)と呼ばれる事情のため、直観的に把握しやすい既存の被差別的属性の選言的列挙は、現時点において保護されるニーズを示唆する政治的無力性の有無・程度と当該語彙で把握される現実の社会集団との関連性を欠く問題を、往々にして生じさせることとなる。より実践的な観点から見ても、個別の被害者の(例えば証言内容を公にすると侮蔑的な視線に晒されるなどの故の)語りの脆弱性あるいは「証言の飲み込み testimony smothering」と、その(公のアーカイブにする営為から離れば忘却されるが故の)語りの不在・消滅に、「記憶」をめぐるポリティクスは挟撃されているのである。「証言の飲み込み」については、クリスティ・ドットソン (2022) 「認識的暴力を突き止め、声を封殺する実践を突き止める」『分析フェミニズム基本論文集』(慶應義塾大学出版会、木下・渡辺・飯塚・小草 訳) の特に217頁以下を参照せよ。
- ²⁷ アイリス・マリオン・ヤング (2020) 『正義と差異の政治』(法政大学出版局) 58頁
- ²⁸ 松宮孝明 (2018) 『刑法各論講義』など、法人への侮蔑罪について、法人に属する構成員への侮蔑となる場合と、単に法人のみへの侮蔑にあたるものとで成立可否を分ける見解が、ここでは想起されよう。
- ²⁹ 前掲注7)『バッド・ランゲージ』ではまさに「レイブ」概念の概念工学的性が論じられている。関連はするものより複雑な社会構造を背景とする課題として、朴裕河『歴史と向き合う』(2022年)の第3章で取り上げられている「慰安婦」の国境による分断事例が挙げられる。本稿と関連する限りで要約すれば、同書ではまさに具体的被害と直結した「慰安婦」の問題が、(歴史的に偶然的なものでありながらも2000年の女性国際戦犯法廷に至る過程で)「戦争犯罪」の枠組みで理解されていく中で、被害者の属性もまた「交戦国」か否かという図式の下に包摂されていくこととなった問題が指摘される。すなわち、「戦争犯罪」という語りは、同時にこの問題を帝国/植民地問題として語る語り口を阻むこととなったことがある。この結果として帝国/植民地に属していた個々の被害者同士(朝鮮人・台湾人・日本人を含む)は分断され、「語り」の紐帯と連帯とを失わせることとなる、というのである。無論、この問題は「慰安婦」問題に限定されることはない。責任帰属をめぐる議論は一定の落ち着きを見せつつあるがいわゆる「徴用工」問題でも、この構図が、臣民動員の旗印の下、実態としては社会における低階層や家庭内でも次男以下など「弱者」徴用=階級動員として反復されることとなったと評価されよう。
- ³⁰ 前掲注20) 林志弦 (2022) 第5章の[追放・抑留と加害の忘却]の項(173頁以下)で詳述されている。それによれば、例えばドイツでは、ホロコーストの加害者としての自己イメージが広く共有される一方で、多くのドイツ人こそナチの第一の犠牲者であるのだとする捉え方も根強く、とりわけ東欧からの「被追放者」としての集団(1200万人が追放され、うち50万人から200万人が引揚げ時に死亡したとされる。)にはその苦難の歴史をホロコーストに連なるものとして持つ者たちも少なくないとされる。事実、追放を受けたドイツ人に対する(ポーランドおよびチェコにおける)抑留収容・虐殺の悲劇は、加害者集団に属するとされた無名の者たちに対する暴力の無制限さを可視化するものであり、上記見解も故なきものではないかもしれない。日本における対応物としては、自分たちを軍部に騙された犠牲者だとする認識と、GHQのオリエンタリズム(=「日本国民は

権威に服従する封建的な慣習の奴隷」であるとする考え)との共犯関係を筆頭に、関東大空襲をはじめとする民間人への空爆や二つの原爆投下による犠牲(そしてシベリア抑留)を頂点としてホロコーストの悲劇と重ねられて理解されてきたことが挙げられよう。これについては、ドイツでは早々に絶版となっていたヴィクトール・フランクルの「夜と霧」が日本語翻訳でのヒットを経て、世界的なベストセラー化を果たしたという過程も目に付くところである。現在、「ホロコースト」が虐殺一般を指すようになりつつある「グローバルな記憶」の濫用を避け、かの一回的な出来事としてのユダヤ人虐殺については「シヨア」という別名を用いる傾向が強まっているのも無関係ではなからう。以上が林の見立ての概要である。

なお、福岡良明「断絶」の風化と脱歴史化 メディア文化における「継承」の欲望(2016)が参考になるとともに、同氏による「記憶の「封印」と「発明」-「継承」と「忘却」の歴史社会学」(2016)に挙げられている知覧特攻平和会館(1975年設立)とその後の「語り口」の変化にも、同様の構造が見てとれよう。同論考では特攻隊基地を有していた沖縄・知覧町における特攻隊員らについての「語り」の持つ歴史的な「歪み」が検証されている。そもそも特攻隊員らは全国から集められたのであり、知覧町の外部から来た者たち、すなわち知覧町民とは系譜的な関わりは全くない。実際、この外部者によって知覧町の主要産業である茶畑は接収され、経済的にも打撃を受けたとされる。加えてその後、終戦直前における激烈な沖縄戦(及び他ならぬ日本軍の手による惨殺)の記憶からすれば、単なる外部者というよりも更に敵対的な位置を、知覧町民の記憶に刻んでもおかしななかった。しかし終戦後、占領下における本土復帰要求の高まりにあつては、特攻隊員たちを含め、自分たちの内部に属するものとして遇される。その後、特攻隊基地跡地には知覧特攻平和会館(1975年)が作られることになるが、それは公的には、特攻隊の記憶が知覧町住民固有の「語り」に取って代わった瞬間でもある。噛み潰された、屈託ある、有象無象の情念が、ここでは覆い隠される。公的に他者の記憶を継承することで自らの記憶を積極的に忘却へと追いやってしまうのであり、いわば知覧の記念館は、その歴史と地理の地場から畏怖を消し去り「封印」するとともに、公的記憶を発明した。以上が福岡の見立てである。

³¹ Th. Nagel(1998), "Concealment and Exposure", *Philosophy & Public Affairs*, 27(1)及び同論考への解説を含む永石尚也(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 第20集』(信山社)第2節(特に「顔」についての言及は129頁)を参照せよ。

³² 米国BLM運動でのこの現れにつき脚注91)を参照。なお、プライバシー侵害をなす発話等、構造的な不可逆性が顕著であるものは別に論じうる。しかし、プライバシーについてさえも、何が知られることを望まないプライバシーに属する事項であるかは当該社会の状況に即して流動的である。また社会において劣遇されている者の場合、別の経済的・社会的利益との関連でプライバシーを自ら放棄させる状況もあり、その確保は一枚岩ではない。前掲注)31永石(2020)第2節・第3節を参照せよ。

³³ 例えばこのような誤用の例としては、2022年に始まるウクライナ・ロシア戦争初期において、「ウクライナ政府によって虐待されてきた人々を保護する」(2022年2月24日のプーチンによる軍事作戦開始演説より引用。)ため、すなわちウクライナによる国家的ヘイトクライムを告発することで成り立つ名目が、ロシアによる侵攻正当化の理由として掲げられていたことが記憶に新しい。こうした名目的使用を許す場合、「虐待」という語彙を用いて差別的構造を発見し、構造を是正する機能は弱体化せられることとなり、特に国家作用としての規制実践へと結びつける根拠としての適格を欠くに至るだろう。「ヘイトスピーチ」についても類似の議論が当てはまる。ホロコーストを典型とする地球規模の記憶文化は、「遠く離れた他者の痛みに共感する道徳的感受性を強める契機」となりつつも、「集合的記憶の民族主義的な競争を地球規模に拡大し、深刻化させる」と指摘されることもある。前掲注)20)林の補論部分(362頁以下)からは、まさにホロコーストの裏面であるヘイトスピーチも上記と相同的な構造を持つことが示されるだろう。

³⁴ Erving Goffman(1981), *Forms of Talk* 第2章[Response Cries]ないし第3章[Footings]においては、対面的な相互行為の分析の中、発話と受話(参与の身分 participation status)との関連が整理されている。まず、①発話産出段階においては、発話主であるところの animator、発話内容の意味・表現を選択した author、発話帰属点としての principle、言及において登場する figure というフォーマットが前置される。次いで、②受話(参与の身分 participation status)段階においては、そのバリエーションとして、直接の聞き手である recipient、発話に向けられた対象である addressee、集まりへの参加が公式に認められた ratified participant と認められていない漏れ聞きを意図して行う eavesdropper、たまたま聞く機会を得てしまった overhearer とそのうち発話者側から把握可能な bystander といった身分が列挙されている。これらが会話の中で前提とされるとともに、続く非難や弁明の中でそれを改訂していくことで会話において適切な発話の条件を更新していくという相互行為の描像は、David Lewisが1979年に“Scorekeeping in Language Game”において提示した会話の Scorekeeping のアイデアにも適合するものであるものといえよう。なお、同書に出てくる諸例は、必ずしも対面的な相互行為に限定されておらず、テレビやラジオなど遠隔性、時差、虚構性をも伴った発話の効果の分析も含まれる。すなわち、発話の時点・文脈等を攪乱させる、発話内行為・発話媒介

行為段階と受話段階の同時性が成り立たない事態をも射程に入れており、理論的にも、上述した発話産出・受話過程の分析でも単に会話場面における適切性を超えて、参与者たちのアイデンティティや続く行為、文化へと接続されることを想定して分析がなされている。この点で、本文に述べた本稿の関心に沿う。

³⁵ 前掲注7)「パッドランゲージ」第6章[6-5節]159頁における蔑称の「禁止説」を参照せよ。なお禁止説の難点として、害に対する禁止の理論的先行という理論上の難点が挙げられる。しかし、この指摘に対しては同説を一般化し、ネガティブな効果を生み出す歴史を持つ動的性質を持つものとして捉えることで解消される。さらに、後述するように法の作用は、この動的性質の上で禁止の限定(単なる悪い表現と禁止されるべき蔑称との区別)と制裁の限定(蔑称の中で、制裁が課されるべきものとそうでないものとの区別)を理論化する営みである以上、本稿の以後の展開上は問題とならない。

³⁶ 江口聡(2022)「分析フェミニスト哲学者によるポルノグラフィ批判(悪いポルノ、悪い哲学)」応用哲学会第14回年次研究大会報告資料を参照せよ。本文中で述べた蔑称使用の受け取られ方に類似した議論として、性的に露骨な表現(ポルノグラフィ性)を持つ作品鑑賞の経験が挙げられる。同ジャンルに属する作品は、特に性差別的・暴力的なそれであってさえも(性差を問わず)一方的な被抑圧の立場を投影することなく鑑賞される。こうした鑑賞態度を、オンラインヘイトスピーチとの比較として検討の俎上に挙げられよう。

³⁷ 例えば、ラジオでのフィクション作品の朗読劇は、たとえ一連の発話が報道調でなされ、たとえそのように広く誤解されようとも報道(誤報)とはならないし、あるいはフィクションならざるテレビのコメディアンが発話がいかにこちら側に語りかけているように思えてもそれはフレームが共有された会話ではない。テレビというメディアを介したそれらの発話は、画面内出来事フレームの下にあり、画面を跨いで存在するのではない。これは、Goffman(1981), Forms of Talk で提示された例であり、実際にラジオを通じて集団パニック現象を引き起こした「宇宙戦争」事件などの事例を基礎としている。他、劇中に登場した蔑称は、フィクション内において蔑称として機能するし、フィクションを理解する上でそう機能しなくてはならないが、現実の聴衆に対する蔑称使用ではない。テレビのコメディアンが用いた蔑称は有限の電波を用いて流すべきではない蔑称使用とはなりえても、視聴者への蔑称使用ではない。ましてや、画面のこちら側のお茶の間からテレビ画面の中のコメントーターにツッコミをいれたところで、せいぜいお茶の間にいる別の人とのフレームにおいてのみ効力を発する(もしも部屋に一人ならば孤独な擬似的会話に終わる)にすぎない。

³⁸ 大屋雄裕(2020)「赦しと忘却」アステイオン93号110-125頁を参照せよ。そこで大屋は、キャンセルカルチャーの一つと目されるリンカーン像をめぐるある騒動に、物理的な抹消を介した記憶と赦しのダイナミズムを完全に失った「責任なき永遠の静寂」を見ている。「ワシントンDCにある解放記念像——白人大統領であるリンカーンの足下に黒人奴隷が跪いている彫像をめぐる騒動をとらえた動画も、議論のさなかで注目を集めていた(https://twitter.com/_sagnikbasu/status/1276682947895197697以下のツイート)。「なぜそんな像を守るのか」と金切り声で叫ぶ女性のBLM運動家に対し、初老にさしかかった頃だろうか、向き合っている黒人男性が静かに問いかけを繰り返している——「誰が払ったのかね」「誰があの像のために資金を出したのだね」。よく知られているとおり、この像が描くのは南軍の首都リッチモンドの陥落後、同市を訪れたリンカーンを奴隷解放に感謝するために取り囲んだ黒人たちの一人が跪いたところ、「私ではなく神に対してのみ跪きなさい」と説いて立ち上がらせたエピソードである。この彫像もまた、解放された元奴隷たちによる献金で築かれたという過去を持つ。歴史的な文脈のなかに置けばむしろ人種解放と平等の象徴であるものを、独立の存在として瞬間的に見ることにより、まったく逆の意味のもとで理解してしまったということになるのではないだろうか。このエピソードは、キャンセルカルチャーの基礎と帰結を象徴的に示しているだろう。すなわちその基礎は記憶の破壊と無知であり、帰結は無反省な自己中心主義である。たとえあるシンボルが文脈抜きに見られた場合に一定の危険性を孕んでいるとしても(ジェファーソンやグラントはその例かもしれない)、その対策としてはBLMが目指すような情報を減らす解決ないし破壊ではなく、情報を増やす解決——そのシンボルが持つ意味を歴史的な文脈のなかで示し我々の反省を促すような注記を添えることが考えられるはずだ。それはまた、情報が適切に与えられれば人々(の多く)はその適切な評価が行なえるだろうという人民への信頼に基づいている。それを拒絶するキャンセルカルチャーは自己の判断力を人民一般に対する高みに置いていることになるだろうがその足下を支えるものはきわめて怪しく、だからこそとすべきか、自己の信念に背く可能性のあるもの、反省を強いる可能性があるものを根こそぎに破壊することが解決策として模索されるのだ」。オンライン空間もまた、歴史的な文脈を無化し、全ての表現がフラットに並べられたかのように誤認されつつ、絶えず物理的摩擦と忘却、そして同時代的なキャンセルの圧力に晒されているという意味で脆弱なものである。本稿末尾においてこの論点をあらためて取り上げる。

³⁹ 折しも2023年初頭、広島市教育委員会による教育教材から削除されたことで話題となった『はだしのゲン』は、(削除の可否自体が法(哲)学的にも興味深い話題を多数含むもの)まずそのメディア特性として商業漫画作品の形式をとる点に着目

きる。教育目的の教材と商業フィクション漫画とは一般的に区別され、教育現場・図書館実務では後者を基本的には排除してきた中で、同書が例外的に配架・使用が許容されてきた経緯は、近時の同書の教育過程からの排除の当否を議論する上でも前提とはなろう。その上で、同作の教育教材からの削除措置に対する肯定的・否定的な反応において（動機はともあれ）焦点化された対象が、方や反戦漫画としての位置づけから歴史の忘却につながるために削除はすべきでないとし、方や専ら戯画的・誇張的な表現方法や侮蔑的・差別的表現あるいは「ヒロシマ」差別への接続の懸念を問題視し削除もやむなしとするなど、論点及びその評価においてすれ違いを見せていた。現下のいわゆる「処理水」をめぐる風評被害及びその報道でも、現在の特定地域への差別的視線を強化する暴力／現在の不安を沈黙化させる暴力としてのフレームのすれ違いを見てとれよう。

⁴⁰ フィクション作品における命題の真偽を論じる基礎として、デイヴィッド・ルイス（1995）「フィクションにおける真理」（樋口えり子訳）を参照せよ。なお、関連して野上志学『デイヴィッド・ルイスの哲学 ―なぜ世界は複数存在するのか』（青土社）が参考となる。戸田山和久（2016）『恐怖の哲学 ホラーで人間を読む』NHK出版を参照せよ。例えば娯楽作品としてのホラー映画作品を鑑賞する際には、まさに恐怖感を求めて楽しむわけだが、ここでの「恐怖感」の意味が現実の「危機感」などとは質的に異なるのはなぜかと美学的・認知心理学的な議論から問うことができる。フィクションを介して「恐怖感」を避けようのない（ある意味では自分に責任のない事態）として安全に享受することは、ある感覚や行動に通じるシミュレーションでもある。その多層化したメカニズム自体が、ボルノグラフィを筆頭として議論される、いわゆる「低価値表現」一般との関係でも、単純化を許さない構造把握を要請するものと思われる。

⁴¹ 孤独な検索行為に勤しむ際、原理的にはあなたの活動は追跡可能であるが、あなたの経験的には窃視的である錯覚を与えるかもしれない。SNSで「要領を得ない」意見を見かけても、それはあなた宛のものではない。仮に引用リツイートやリプライを受けた際にでさえ、そうでありうるのである。これらは、ブログやSNS等の発信・受信の形態と使用のされ方が構築してきた諸フレームの混線が起こっていることを示唆するだろう。このことからすれば、オンライン空間で蔑称が使われる際にも、その語彙によって生じる効果はフレームの誤解によるものであることが大かかもしれない。そうだとすれば、蔑称や侮蔑的・差別的発話を専らその語彙（効果）にのみ着目し、媒介するメディア効果抜きに抽象化して論じることは、当該発話の機能を見誤ることにつながるだろう。とりわけオンライン上のヘイトスピーチについては、後述する拡散性や永続性という性質から、より広範な「害をなす」発話とともに、それへの自動検知やプラットフォームによるコンテンツモデレーション等、手法面・主体面での拡張もなされる。これは同時に従来の理解を主体面・作用面において拡張した準・規制による新たなリスクをも生じさせるだろう。複数の「メディア」を介し、法規制もあるいは社会的非難を含む事実上の規制作用もなされるのが通常であるオンライン空間においては、上記の文脈読解にあたり、当該メディアの技術的・社会的特性を踏まえた検討が要されるのである。

⁴² ある規制が対象とした（例えば反ムスリムの）ヘイトスピーチへの規制によって、別の保護されるべき集団の自由を逆に制約したり、まさに保護されるべき集団への別の害を生み出すことに留意が必要であることは、前掲注17）で人種差別撤廃委員会も注記している通りである。

⁴³ 齊藤愛（2015）『異質性社会における「個人」の尊重』を参照せよ。表現の自由の保護範囲論の枠内で集団侮蔑を内容とする表現使用を論じる憲法学上の議論もまた上記文脈を構成する要素として包含される。次稿にて詳述する。

⁴⁴ 横浜地裁川崎市部決平成28年6月2日など、リアル空間における本邦の事例群の多くで見られるヘイトデモ型については、個人の人格権の一部に含まれる平穏に生活する権利、自由に活動する権利、名誉信用を保有する権利等、広範な利益を民事上の損害として認め、差止めを認めている。他方で上記デモがなされた川崎市でその後に制定された川崎市差別のないまち人権条例（2019年制定）ではヘイトデモについての罰則（第12条）が付されているが、あくまでも公共の場において面前でなされるものに限定され、罰則に至るまで慎重な手続きも規定している。このような限定に現れているように、刑事法領域において、一般的に平穏な生活環境を保持する利益や自由に活動する利益が（刑事的制裁を背景として実現されるべき価値として）直接の保護対象とされるかについては疑わしいものと考えられる。

⁴⁵ 例えば、2018年、オンライン上にブログ記事を投稿し、在日コリアンの少年に対する誹謗中傷が侮蔑罪に当たるとされ、科料9000円の略式命令に至った事件や、脅迫の嫌疑での書類送検後、神奈川県迷惑防止条例違反（つきまとい行為禁止）で罰金30万円の略式命令が出されたものなどが報道されている。

⁴⁶ 典型として、京都朝鮮学校公園占用抗議事件（最終2012年2月23日）がある。

⁴⁷ なお先述したゴフマンが述べている例のように、どちらかといえば、自室のテレビ画面の前で悪態をつく、フレームの誤作動状況に近いかもしれない。この悪態をついているものの発話が意図的に発話者に届く場合を別として、偶発的に漏れたとしても、それ自体は、流通の問題である。

⁴⁸ Langton(2018)，“The Authority of Hate Speech” の分類が参考になる。なお、本邦においても重なる分類として提案されている

ものとして直接的・間接的なヘイトスピーチを分けるものがある。那須祐治 (2019) 『ヘイトスピーチの比較研究』 487-493 頁を参照せよ。

⁴⁹ オノラ・オニール (2016) 『正義の境界』 (みすず書房、神島裕子 訳) 第 5 章でいうところの、重要な利益とバスターにされた選択の自律性に関する「(あなたが) 拒否できない申し出」の問題である。

⁵⁰ 無論、こうした人権制約局面における条例制定との関連では、そもそも青少年保護というトピックが (特に法律の留保からの検討を含め) こうした分権性に適った対象であるのかについては、改めて検討する余地はあるだろう。もしもこの検討によって、当該トピックが本来的に国家に属すべき事項であった (にもかかわらず地方にいわば「投げられて」きた) ことが判明するならば、然るべき審議手続きを踏まえた国家の判断を要求することになろうし、反対に (本文に述べたような全域性に由来する影響関係の拡大から) 当該トピックが国家にも地方にも属すべき事項でないことが判明するならば、そもそも法的介入を控えるべきことに帰結しよう。この点について、大屋雄裕による講演 (2022) 「“危険” な科学・技術に関する情報の統制と自由を考える」 (2022 年 9 月 24 日) が参考になった。

⁵¹ 典型としての青少年保護のためのサイトブロッキングに加え、近時話題となった著作権保護のための「漫画村」サイトブロッキング、あるいは侮蔑罪の 2022 年の厳罰化に、空間特性に応じた介入の拡張とその条件設定のバリエーションを見ることができると。

⁵² 前掲注 45) 「川崎市差別のない人権尊重のまちづくり条例」では、インターネット上のコンテンツに対する拡散防止措置 (プロバイダ等への削除要請) が定められているが、そこでは該当する投稿が川崎市内でなされたものあるいは川崎市民を対象とする場合などの限定が付されており、また投稿に対しては禁止・罰則は付されておらず、審議会の諮問を経たのちのプロバイダへの削除依頼手続き及び公表措置のみが規定されている。上記の権限との調整の結果として把握されよう。同条例 FAQ ページを参照せよ。 <https://www.city.kawasaki.jp/templates/faq/250/0000128378.html>

⁵³ Joel Feinberg(1984), *Offense to Others (The Moral Limits of the Criminal Law)* の第 8 章 [Mediating the offence principle] を参照せよ。

⁵⁴ いわゆる「囚われの聴衆」に関する拒絶困難性を構成する事情においては次の事情、すなわち文字情報はもとより画像情報でさえ強制的に耳に入ってくる音声等と比して態様としては一瞥の後に破棄することが可能であることや、その発信者が公的地位を代表するものであること (政府言論性)、または物理的閉域 (あるいは移動の目的のために相当時間の物理的拘束を強いられる公共交通機関車内や生活の糧を得るための職場などそれに類した空間) における強制的・反復的表示機能により避けることができないことといった事情が考慮されることを、ここで思い起こすことができるかもしれない。この点につき、横大道聡 (2013) 『現代国家における表現の自由一言論市場への国家の積極的関与とその憲法的統制』 第 11 章 [政府言論の憲法的統制] 及びその前後の関連する章を参照せよ。「囚われの聴衆」法理の適用場面は、私人の「表現からの自由」を守るための政府の規制 (利益) の正当化場面である。そうである以上、「囚われの聴衆」法理は、例えばその発信者の地位 (政府言論性)、場所 (物理的閉域性)、態様 (非文書=音声性) 等により自ずと限定される。自ら検索して、あるいは自ら構築した TL を通じて特定の発話 (文字情報) が視界に入ることは、あなたと Google との、あるいはあなたと Twitter との共同制作物である。この結果として、オンライン空間において情報の高度の流通性が仮に存在するとしても、ジェレミー・ウォルドロンという「社会的環境の汚染」、すなわち尊厳あるいは包摂性という公共財に対する計算された攻撃の量的拡大とみる見解は、表現のオープンさをパブリックな影響関係へと即座に連結している点で妥当ではない。

⁵⁵ 名誉毀損における「社会的評価低下」判断にあたっては、紙の上に固定化・活字化された新聞・雑誌等の文章メディアについては「一般読者の普通の注意と読み方」基準が、音声・映像情報が次々と流れていくテレビ放送については「内容・印象等の総合考慮」基準が適用されてきたことが、ここで参考になる。特に後者については判例上、「録画等の特別の方法を講じない限り、提供された情報の意味内容を十分に検討したり、再確認したりすることができない」特性が考慮されていたことは、オンライン上の表現の解釈に当たっても参考になる。以上につき、村田健介「インターネット上の名誉毀損と民法法理」法学セミナー 66 巻 12 号 (2021) を参照せよ。

⁵⁶ すなわち、参加者はオンライン空間において別々の「社会的現実」あるいはネルソン・グッドマン (2008) 『世界制作の方法』 (ちくま学芸文庫、菅野 盾樹・中村 雅之 訳) の言葉を借りれば「世界」のヴァージョンを制作し (その適合性評価を行い) ながら、しかしそれぞれが同じ世界を共有しているかの想定を構築する側面を持つという擬似同期性を指摘できる。この点につき、動画共有サイト・ニコニコ動画のコメント機能に即した擬似同期性の分析として、濱野智史『アーキテクチャの生態系』及び村上裕一『ゴーストの条件』第二部第四章・第六章の分析から示唆を得られるかもしれない。ヘイトスピーチの蔓延状況を網羅的な機械的走査から把握しようとする試みは、社会的な言説を取り出す点で有益ではあるが、少なくともここで問題となって

いる各人の認識を介して生じる害を明らかにすることは異なっている。ここでの関心は、世界の中にある言説一般の悪性を除去することではなく、社会的状況の支援を受けつつ、オンライン空間の特性の中、個別的な認識の上で現出する害への規制の発動条件に向けられている。

⁵⁷ これはいわゆる Twitter における「エアリブ」を想起すれば理解しやすいだろう。SNS に投稿する場面では、想定された相手が投稿を見るだろう予期は、相手が現にその投稿を目にする確証はなく、必ずしも実現しない。また実際に目にしたとしても、その予期通りに実現したか否かは、相手からの直接的・間接的反応によってのみ事後的にのみ確認しうる。こうした予期の即時的実現と確認の明示性が確保された対面状況との差異がオンライン空間では顕著なものとなるだろう。この点でゴフマン (2023 年)『日常生活における自己呈示』(ちくま学芸文庫) (原著 1959 年) の第 3 章「領域とそこでの行動」は特に示唆に富む。前掲注 37) 等ゴフマンに言及した箇所も参照せよ。

⁵⁸ もちろん、ここにはグラデーションがある。一方の極には、Zoom 面接などが典型であろうがリアル空間における対面コミュニケーションと寸分違わぬコミュニケーションを (技術的にも、あるいは規範的にも) 実現するものがあり、他方の極にはリアル空間とは異なる空間・時間・関係フレームの上で全く別のコミュニケーションを実現するものがある。中間には LINE 等の既読機能などがあるかもしれない。

⁵⁹ 無論、リアル空間においてもよりマイルドな介入として「拒否型」があるが、これはオンラインでは起こりづらい。若林美奈 (2021)「集団的差別的言動と不法行為」は、ゴルフクラブ入会拒否や賃貸物件入居拒否など不平等な取扱いをなす類型を拒否型事例と整理した上で、私的自治が一定の範囲で制約され、その限度を超える場合の効果として契約無効及び人格権侵害による不法行為を導く。さらにこの拒否型と区分される類型として、平穩を突如として乱す (典型的には本人に対して名指しで行われる) 排斥・危害告知・誹謗中傷をなす差別的言動型のヘイトスピーチの性質を取り上げている。この区分自体は妥当であろうが、他方で事実上の公開性と事実上の匿名性 (という想定) に支えられているオンライン上においてはそもそも属性開示が不透明である以上は拒否型は起こりづらく、また平穩を乱す性質も対面状況とは異なることは念頭に置いてよいだろう。

⁶⁰ この性質から、オンライン空間をリアル空間における公共空間と同視したり、あるいはそれを前提としてパブリックフォーラムとしての制約除去根拠として用いる見解は、空間特性の把握において問題を持つ。関連する点を指摘するものとして、土井翼 (2023)「SNS とパブリック・フォーラム論—非物理的パブリック・フォーラム論の無用性」情報通信法学研究会 メディア法分科会 (令和 4 年度第 2 回) 報告資料を参照せよ。なお関連して、土井翼 (2022)「傍聴人による議事の撮影を許可制とする県議会規則及び映画製作会社による撮影許可申請につき個別事情を考慮せずにした不許可処分が合憲かつ適法とされた事例 [那覇地判令和 2 年 8 月 5 日判批]」自治研究 98 号 7 号) も示唆に富む。

⁶¹ もちろん、悪しき例として掛け金を支払わない事による過剰さに辿り着きうることはある種のキャンセルカルチャーの例においてみられる。この点につき、大屋雄裕 (2022)「キャンセル・カルチャーが孕む二つの文脈」Voice (537) 148-154 の「賭け金は何か?」の節を参照。

⁶² Robert M. Simpson “Is Hate Speech Punishable?” (2023)「ヘイトスピーチは罰しうるか?」永石尚也・本田康作 訳、『差別・ヘイトスピーチ・法』(近刊) 第 2 章

⁶³ 前掲注 38) 大屋 (2020) の引用部を参照

⁶⁴ Catharine A. MacKinnon(1996), Only words などが代表的であるが、表現の自由に重きを置く論者の中でも、ナイジェル・ウォーバートン (2015)『「表現の自由」入門』(岩波書店、森村進・森村たまき訳)でも、「ポルノグラフィ」表現について、その暴力との結びつきを理由として、実証的な検証を待たずに規制する可能性が示唆されている。

⁶⁵ 特に「安心」に基づく規制積極論に対しては、ヘイトスピーチの態様やその文脈 (特に差別的慣行が制度的裏付けを有しているのか、あるいはその慣行に対抗する是正作用が制度化されているのか) によって、当然毀損される安心は質的・量的に異なるはずであるにもかかわらず、その存在自体が安心を奪うとするのは、実証的な社会科学的知見を避けるものであるとの批判がなされている。前掲注 64) Robert M. Simpson (2023) (近刊) 第 2 章を参照せよ。

⁶⁶ 林岳彦 (2023)「Evidence Based Practice にとって「良いエビデンス」とは何か」に倣い、エビデンスを文脈特定性や社会的妥当性を超えて過剰利用するケースを念頭に置いてみよう。エビデンスそのものの科学性と政治性が両立するように、エビデンスの科学性とエビデンス使用の政治性とは両立するため、経験科学的研究が参照される場合にはその利用の適正性が鋭く問われる。井頭昌彦『質的研究アプローチの再検討』(勁草書房) 所収の同論考を参照せよ。

⁶⁷ 次稿 (2) 第 2 節・第 3 節で詳細に比較するが、例えば本邦において、法の効果としての責任に着目するならば、特定人を対象とした、具体的な害を伴う差別的言動については、刑事上は名誉毀損罪・侮辱罪を構成するとともに、民事上は不法行為としての責任を問われる。その一方で、不特定多数の集団に対する差別的言動については、法益となる外部的名誉・主観的感情

の不定性と、その認定上の困難・恣意の危険から、本邦においては直接に規制の対象とはされていないし、その民事賠償も困難である（山本敬三（2013）「差別表現・憎悪表現の禁止と民事救済の可能性」国際人権 24号を参照せよ）。この区分は単なる規制の不足を示すものではなく、（法）規制の正当性の問いと正統性の問いに答えた結果でもある。なぜなら、①集団に対する差別的言動については、個別の法益を具体的に侵害するケースとは質的に異なる上、②輪郭が明確かつ限定された個人的法益とは異なり、集団の範囲設定によってはむしろ当該集団にとって不利となる規制をも招くリスクさえあるためである。

⁶⁸ 前掲注 62) Robert M. Simpson (2023) (近刊) 第 2 章所収。なお、この批判は、個別のヘイトスピーチそのものを原因とした害の発生に直接着目するのではなく、社会における既存の不平等が維持・継続される状態へと当該発話がどれだけ寄与しているか、その徴候 (symptom) としてヘイトスピーチを取り扱う提案に根ざしたものである (同第 7 節)。ウィトゲンシュタインに由来する「徴候」は、論理的な必要条件を示すものでも十分条件を示すものでもないものの、因果に関する経験命題の「確証」を促しはする。ヘイトスピーチが害を作り出す (単称的) 因果関係は複雑かつ多重的である一方で、社会における既存の不平等のメカニズムは社会科学的分析に適していることに照らしても、方法論的に見てシンプソンの提案それ自体は一貫したものである。ここから経験科学的な研究によって明らかにされるだろう実際の言語の働きや私たちの持つ価値の上げ下げの実態にシンプソンが期待を寄せ、ウォルドロンを批判するのは自然である一方で、経験科学的な研究が刑罰の執行を留保なく基礎付けうるものと位置付けている点は論争含みである。

⁶⁹ 司法上の判断にも、こうした検抑性を見てとることができる。大阪市ヘイトスピーチ条例公金支出無効確認請求事件 (最判令和 4 年 2 月 15 日) は、ヘイトスピーチの対象については明示的に踏み込むことなく、その制限についても、「制限の態様及び程度においても、事後的に市長による拡散防止措置等の対象となるにとどまる」ものであり、「当該要請等に応じないものに対する制裁はなく」、「認識等公表についても……法的強制力を伴う手段は存在しない」とする。この結果、個々の表現行為・目的 (+ 動機) を定義・分類する困難、認定・該当性判断の困難に対して、審査会設置・運用を含むメタ的なプロセスへと課題を移送することを法的に根拠づけているものと理解できる。なお、判例評釈として、高瀬保守 (2023) 「大阪市ヘイトスピーチへの対処に関する条例 (平成 28 年大阪市条例第 1 号) 2 条、5 条～10 条と憲法 21 条 1 項」、毛利透 (2022) 「大阪市ヘイトスピーチ対処条例の合憲性 / (最高裁三小法廷令和 4 年 2 月 15 日)」、阿部和文 (2022) 「[時論] 最高裁第三小法廷令和 4 年 2 月 15 日判決に関する覚書——大阪市ヘイトスピーチへの対処に関する条例の合憲性に関する判断を中心に」ジュリスト 2022 年 7 月号 (No1573) には初稿執筆後に接した。なお公表の効果については、あわせて「平成 30 年 1 月 17 日 大阪市ヘイトスピーチへの対処に関する条例の施行に関する事項について (答申)」7-8 頁も参照せよ。

⁷⁰ J. S. ミル (1971) 『自由論』(岩波文庫、塩尻公明、木村健康 訳) 169 頁

⁷¹ 同上 47 頁。これを支える社会的な支えとは、「自己自身が多数者の優位に反抗しつつ、公衆の意見や傾向と相違する意見や傾向を保護することに関心を持つ、しっかりとした社会力」(148-149 頁) を意味する。もちろん、ヘイトスピーチ規制の難問は、「多数者の優位への反抗」と「公衆と相違する意見・傾向」の保護が素朴には結びつかず、現象レベルでは主として (例えば日本における日本国籍を持つなど) 特定の属性における「多数者の優位」を象徴的に利用することでなされるものでありながら、同時に社会的な支えというメタレベルでは「公衆の意見や傾向と相違する意見や傾向」にも含まれる個性の発露でもありうるという、一見混乱した帰結を生むためである。次稿第 4 節ではオンラインヘイトスピーチの持つ言語行為上の効果を分析することで、その真理想定の正当化に資するように「社会の支え」を通じた多様性を保持する環境が整えられることで、その特定の類型が許容される条件を明らかにする。

⁷² 若松良樹 (2020) 『醜い自由』(成文堂) 5-6 頁

⁷³ 前掲注 70) ミル (1971) 『自由論』15 頁

⁷⁴ 前掲注 72) 若松良樹 (2020) 153-155 頁では、個人の発展可能性としての個人内多様性と、個性への不寛容を拒否する個人間多様性を双方含むものとして整理されている。そして当然ながら、「多様性が価値を有しているとしても、多様性のために今何をすべきかはそれほど明らかではないのであり、政府が直接に多様性を実現することには限界がある」(191 頁) にも留意する必要がある。

⁷⁵ 前掲注 70) ミル (1971) 『自由論』170 頁

⁷⁶ 同上 15 頁。「社会自らが暴君であるときには (社会を構成している個々人の上に集団としての社会が君臨しているときには) 暴虐遂行の手段は、社会がその政治上の公務員の手によって行いうる行為のみに限られていない。(中略) 社会は自己の命令を自ら執行することができ、また実際に執行しているのである」。保護されるべき属性をその内容において特定する法規制を求める要請は、反転してマジョリティ側の武器として通用するリスクをも持ち、そのリスクは不均衡な形でマイノリティ側にこそ降り掛かりうる。

⁷⁷ ここで、新型コロナウイルス感染防止における各種「要請」や解釈指針の提示、さらに近時議論されているところでは実効性確保手段としての各種ナッジの問題等、本来は法規制に対する統制が働くべき場面を執行側が回避する動きが進んだことを思い起こすことができる。これは国家による社会的暴虐への委任ともいえ、こうした統制からの逸脱事例を正統性や正当性から法領域に包摂することは法の支配のコロラリーともいえよう。こうした非-法的手法が一般的に内包する問題は、本文で述べたような中間領域構成型のヘイトスピーチ解消法が内包する問題とは質的に異なる。

⁷⁸ インターネット上の口コミの許容性判断につき、①それが事実摘示でも意見表明でもない主観的な体験表明であり、②投稿単体ではなくその他投稿を含む総体において社会的評価を左右するとの性質に着目し、従来の新聞、雑誌、テレビ等の伝統的媒体についての基準とは異なる判断枠組みとして受忍限度を用いているものと分析するものとして、内田貴 (2023)「インターネット上の口コミの削除請求——その法律構成について」ジュリスト 1586 号がある。他方、こうした口コミの被害を訴える側の当事者の属性として、同稿は、③ (医師・弁護士等の専門職が原告となるケースが比較的多いことから) 不満をコントロールすることが求められる専門職的な業務特性や、④事業者と消費者との間の情報の非対称の解消、⑤事業者による対抗的な情報伝播能力にも着目し、受忍限度という枠組みが妥当する場面を理解しようと試みている。これらの③④⑤は多くのオンラインヘイトスピーチ事例においては直接は当てはまらず、別の要素によって補完される必要があるだろう。

⁷⁹ 例えばヘイトスピーチ解消法が専ら指針提示するに留めつつ、ヘイトスピーチへの市民的対応を促進する趣旨は、法による対処のリスクと法以外の手法を通じた対処によるリスク (そして法以外の手法が法が統制して対処させることによるリスク) との間の緊張関係の中で、特定の表現に対する「より多くの言論」を求めた結果であると、この観点からは解釈できよう。(刑)罰の執行について消極的判断に至ったとしても、それ自体としてはヘイトスピーチ及びそれに準ずる差別的発言への放置・容認を意味しない。そもそもヘイトスピーチ解消法それ自体は、個別の表現行為・表現の良し悪しの判断へと誘導するものでもない。同法は、個別のヘイトスピーチの「悪さ」を基礎づける公的規範というよりは、市民社会のイニシアティブで構成する「空間設計」を志向する公共圏の維持規範であり、同時に、その道徳的過剰を抑止することをとも志向する公共圏維持の規範であると言い換えられる。すなわち、同法は、第一に、同法に反する個別の表現についての判断を事前決定することがないし、同法に明示的に言及する裁判例群も、同法の趣旨に照らして個別法上の不法行為や名誉毀損等の該当性を判断してきたものである。同法における過小包摂があえてなされたものだと考える場合、本邦におけるヘイトスピーチのコアに属するターゲットだけを取り出したことからは、本法における対象限定が終局的なものではなく、常に周辺の差別実践を取り込むべき (明示的には誘導することなく) 議論に開きうることを示唆していると解釈できよう。この結果でもあるが、第二に罰則に裏付けられた規制を欠いており、専ら市民社会における攻防における主軸となる価値を表現している (政府言論) と考える場合には、まさしく市民社会においてときどきに変転する「あるべき表現」と「そうでない表現」を競い合わせる空間が維持されているか否か、典型的には言論空間における不平等性・競争排除といったメタ的な「悪さ」が現出しているか否か、という二階の「悪さ」に焦点を当てているものとして、罰則の不在を積極的に理解できる。これら二点を通じた、非規制的で自己抑制性を (ある程度) 備えた競争環境維持装置として、同法の限定性は (先に挙げたオーストラリアの 1975 年連邦人種差別法 18c (1995 年刑事罰導入) における慎重な手続・運用状況と同様に) 積極的に把握できる。

翻って、法という特殊な手段で言論空間における競争を上から内容的に改善することは、各種の困難を伴う。前述した権力の暴走抑止はもとよりとして、より基底的には、構造的な不平等・差別による脆弱な発話者・発話内容へのエンパワメントによる改善を抜きに真实性・公共性の判定によってこれをなすことは、却って公共空間における言説空間が目指す共通価値 (自尊感情と社会的承認の価値) を毀損することが難点となる。次稿第 4 節を参照せよ。

⁸⁰ 前掲注 38) 大屋雄裕 (2020)「赦しと忘却」アステイオン 93 号 110-125 頁

⁸¹ 表現の自由の行使の結果として排除を正当に要求できるのは、「私のうち (=私的空間) でやるな」であるに過ぎず、「(お前の私的空間の) うちに閉じこめれ」と公共空間からの撤退を強いることではない。

⁸² J. S. ミル (1971)『自由論』(岩波文庫、関口正司 訳) 37-38 頁参照。そこでミルは、次のように述べ、主張の真理性よりも主張の有用性を優先する見解の自己論駁性を指摘している。「信条の中には、幸福にとって、必要不可欠とまでは言わないにしても非常に有益な信条があるのだから、それを支えることは、他の社会的利益を保護することと同様に政府の義務だ、と主張されるのである。こういう必要性があり、政府の義務とするのにこれほどふさわしい場合には、政府は、無謬だとまで言い張らなくても、大方の人々に支持されている意見にもとづいて行動してよいのであり、それどころか、そのように行動しなければならぬ、とされる。また、有益な信条を弱体化させたがるのは悪人だけだと主張されることは多いし、それにもまして多くの場合、[口には出さなくても] そう思われている。だから、悪人を規制し、こういう連中だけがやりたがることを禁止するのは何も間違っていない、という考えに至るわけである。このような考え方では、討論への規制の正当化は、主張の真理性の問題

ではなく、主張の有用性の問題になる。これで、意見に対する無謬の判定者だと主張する責任を回避する手段ができたとして自己満足するのである。しかし、このように自己満足している人々は、無謬性の想定が、一つの点から別の点に移動しただけだということに気づいていない。」

⁸³ 前掲注 70) 35 頁でミルは、端的に「判断を信頼できるのは、訂正する手段がいつでも手元に保たれている場合に限られる」と述べる。なお、第二章冒頭において、意見表明を沈黙させることに弊害として、現代代のみならず後世の人々にとっての不利益、その意見に賛成する人々のみならずその意見に反対する人々にとっての不利益についても検討している。

⁸⁴ ジェレミー・ウォルドロン『ヘイトスピーチという危害』第 5 章 (126-128 頁) を参照せよ、

⁸⁵ アヴィシャイ・マルガリート (2017) 『品位ある社会〈正義の理論〉から〈尊重の物語〉へ』第 7 章を参照せよ。もしも主観的に価値づけているものであれば、アイデンティティ構成的な属性はその自律的選択により、客観的には脱ぎ捨て難くなり、もしも社会的に押し付けられたものであれば、その差別的環境のために主体的に脱くことが困難であろう。ウォルドロンが理念的な平等な尊厳によって等閑視するのは、このアイデンティティ構成における主観的であるとともに客観的な側面である。アヴィシャイ・マルガリートは、ウォルドロンのいう感覚・感情をさらに区分し、ウォルドロンのいう尊厳へと結びつけている。そこでは社会的な名誉を毀損する侮蔑と自尊心を傷つける屈辱とは区別されるが、後者の屈辱は単なる主観的な不快にとどまらない。屈辱は、共同体に属する人間存在を共同体から拒絶すること、すなわち人間存在を普通の人間以下の存在として取り扱う事態に由来するが、そこで自尊心は、自分が傷ついているか否かの主観的感覚にかかわらず傷つく。なぜなら、自尊心の傷つきは、単に孤独のうちにはなく、他者が自分自身をどのように取り扱っているかを無視することはできないためである。ウォルドロンを引き継いでいえば、上記感覚・感情の中でもとりわけ自尊心だけは「政治的狀況」のうちにおいて、それを支える基底としての地位と不可分の共同体感覚を示すのである。

⁸⁶ 前掲注 84) 159-162 頁では、これを「アイデンティティの政治の無責任さ」とし、「ある争点が私のアイデンティティにとって決定的に重要だということによって、私はその争点についての私の見解を政治的に交渉不可能なものとして提示する」ことによる集合的決定の挫折と呼ぶ。

⁸⁷ ミルを引きつつ検討した通り、重要となるのは、公私の間にあり、空間的・時間的に厚みを持った、上記の受忍を支え、寛容を育む「社会的支え」であったことが想起されよう。すなわち、個人的な関係解消や説得・批判と全般的な不可視化・沈黙化との間にある、「社会の支え」による抵抗の潜在的可能性を常に保存し、反対に機会主義的な否認を許さない議論継続を承認し合うための中間領域こそ、多様性の観点から見た個人・社会の発展という目標の基層にあったミルの考えであった。

⁸⁸ 前掲注 70) ミルは第二章末尾で、誹謗、詭弁、論点の隠蔽、はぐらかし、議論の歪曲といった「節度」を欠いた主張について述べている。「節度」を欠いた主張は、確かに道徳的に非難されるべきであるが、他方でこの要求に従い相手を討議から退出させることはできないし、それは法及び権力の任務ではない旨、明確に主張している。その理由としては、その「節度」要求が概して少数者に対して向けられがちであることを挙げている。

⁸⁹ 本文での「ブロック」は、隠語等の使用による「含み」を解除する Rae Langton(2018), Blocking as Counter-Speech の用語法に従う。ここで 1.1 節に立ち返り、いま一度、蔑称使用・隠語使用による害との比較を行うことは有益であるだろう。蔑称は、発話者が意識せずともそれを使用することによって社会的現実あるいは世界をカテゴリカルに切り取り、「制作」する側面とともに、意識的にそれを使用することによって、自らがその語を取り巻く社会的状況においてどの位置にあるのかを周囲に顯示する側面を持つ。もちろんこの二側面は、他の用語（例えば、一見したところ中立的な地理記述の用語である「中東」や「極東」に含まれる権力性など）や慣用語でも一般的にそう機能する可能性はある。その一方で、語に対する訂正過程や語に含まれる脱権力化プロセスを介して、語を自分たちの手元に取り戻す抵抗の動きもまた組成しうるのである。この蔑称の二側面に対して、隠語はその後者にあたる顯示的側面につき、弁明として「そういうつもりはなかった」という否認の余地を持つ。蔑称については「そういうつもりはなかった」としても特定文脈の下で現にそう使用したことが非難されうるのに対し、隠語はその使用の文脈をより特定状況依存的な形でもち、それゆえに「そういうつもり」の意図が弁明として機能する。例えば、相手の責任に属する危害や相手が信頼に値しないことを仄かす隠語や不安の表明は、非認知的な影響力を行使しながら、事後的にその行使意図を否認できる余地を持つ。いわば隠語使用は、語に対する訂正過程や語に含まれる脱権力化プロセスそのものが脱白させられるリスクを生じさせるのである。即ち、告発された害を発話者自ら「隔離 quarantine」しうるのであり、更に本文 1.2.4 及び前掲注 39 の通り、この戦略選択は主体の（一見したところの加害・被害の）別を問わない。

翻って、オンライン上のヘイトスピーチもまた蔑称・隠語と同様に、その文言上の意味内容として、社会におけるマイノリティ成員に対して、彼らの社会的地位を間接的に「普通」以下に引き下ろすことを示唆／顯示する側面に近い言明ではある（沈黙化や抑圧と同様）。オンラインの特性を踏まえれば個別的・集団的に抵抗が可能であるとしても、繰り返される当該言明と社会

における「寛容」が、具体的な当該成員の社会における地位を現に固定化する「おそれ」には、社会科学的な蓄積を踏まえれば一定の根拠が認められるかもしれない。その上で、隠語特有の性質をも引き継ぎ、オンライン上のヘイトスピーチは上記の抵抗の際に、弁明としての「そういうつもりはなかった」と否認される危険を有するともいえるだろう。しかし、こうした否認の危険が認められるとしても、沈黙化への抵抗、(否認に対する)可視化・再否認(Langtonの言う「ブロック」)も更に可能である。従って、問題は(単純な加害・被害の図式に取まらず)主体の別を問わずに行使される沈黙化、抵抗、否認が、どれほど肯定的な支持やアーキテクチャの支援を受け、各々の発話が容易(あるいは困難)となっているかにある。この支持・支援の社会的・歴史的文脈の特定なしには、沈黙化や抑圧についての哲学的思弁は、ヘイトスピーチの現象的側面の解明という課題を除き、規制の文脈ではほぼ機能を失ってしまう。本文2.2.1で見た通り、ある害を除去する規制にはリスクトレードオフが存在し、規制がなす「悪さ」は、その正当性のみならず、その正統性にも及んでいるためである。

⁹⁰ 志田陽子(2022)「表現の自由」をめぐるワインディングロード」現代思想50巻3号は、まさにこの狭義の意味で、契約当事者外からの当事者の解職要求や社会的信用を低下させるような働きかけもまた「キャンセル」と捉える。本稿もこれに即し、社会的・経済的地位剥奪としての解雇や契約関係解消、オンライン空間からの締め出しという意味ではフォーラムからのアカウント削除や特定属性への封殺・沈黙化等のハードな形でなされる「キャンセル」という意味で、この語を理解する。なお、こうした立場に対しては、「キャンセル」の意味を広義に取ることで許されるキャンセルもそうでないキャンセルもある、と反論されるかもしれない。しかし、こうした概念の混同は(キャンセル現象の記述や例示としては妥当であるとしても)規制をめぐる議論の焦点を逸してしまうだろう。特定の発話や不品位行動に対する非難・批判・説得や、発話者である者との間の友人関係・契約関係からの個別的な離脱といったソフトな「キャンセル」と、そうした行動を他者に要求して集団的に行動することや法を用いた禁止・不可視化措置といったハードな「キャンセル」との間の境界が曖昧化されることは、個別的発話の許容性判断基準の定立を目指す場合には資するところがない。

⁹¹ 前掲注38)大屋雄裕(2020)ではまさしく黒人差別を訴える“BlackLivesMatter”に対して、“AllLivesMatter”やその他訴えが、BLMに対して敵対的で人種差別的であるとされ、キャンセルされる事態が紹介されている。「すでに二〇一六年の段階から、「すべての命は大切だ」(AllLivesMatter)という表現はBLMに敵対的であり人種差別にあたると批判されるようになっており、民主党支持者として知られる女優ジェニファー・ロペス(彼女自身がプエルトリコ出身の両親のあいだに生まれたラティノーである)でさえ、この表現をハッシュタグとして付したツイートの削除と撤回に追い込まれている。二〇二〇年七月には、「すべての命は大切だ」だけでなく警察官(しばしば青い制服を着ている)の生命保護を訴える「青い命も大切だ」(BlueLivesMatter)や、やはりマイノリティグループと目されている「アイルランド系の命は大切だ」(IrishLivesMatter)といったスローガンを付した商品の販売を取り止めることを、アメリカ最大の(そして世界最大の)スーパーマーケット・チェーンであるウォルマートが表明するに至っている」。特定の社会的文脈の下で上記敵対性は確かに発現しうるが、その発現は多重に条件的である。

⁹² 前掲注61)大屋雄裕(2022)を参照せよ。もちろん、私的自治領域における一対一の契約内容上の拘束や契約解消が、その効果として何らかの地位からの離脱や利益への期待の剥奪を伴うものであったとしても、違法な拘束・解消あるいはそもそも優越的地位の濫用等による例外を除き、個別的な関係からの離脱や期待剥奪がここでいう「キャンセル」の作用を果たすものとは見做されないだろう。しかしそこから更に踏み込み、当事者の契約には書かれていないある観点からの「不快」あるいは不品位性等の道徳性を理由として、当事者外の者が(批判を超えて)外部から圧力をかけることは、上記の意味での通常の法的な意味での私的自治の範疇の限界点に位置していることもまた明らかである。なぜなら、契約関係であればそれによる履行請求や解消の妥当性、あるいはその地位上の格差(例えば優越的地位の濫用法理の適用可否)が、法的な意味で終局的には対等な道具立ての下での解決を志向する司法的解決に付されることが期待されるのに対し、外部からの匿名的圧力の集積は、自らの利害に直接には関わらない場面において(例えば当該人物と雇用先との契約主体を介して)「キャンセル」の結果を間接的に引き起こしうるにもかかわらず、その(比喩的には法の下に出頭すべき)責任主体性を曖昧化したまま、当該結果に対しての責任を取ることを避け、あるいは結果をめぐる回復の営みに参与せず、事後的に自らの行為と結果との間の連関を機会主義的に「否認」しうるためである。

比較対象として、例えば消費者集団が自らもまたその商品の利益を享受する主体であることを理由とし、その自律的意思を集積することを呼びかける不買運動や、社会的な誉れを総合的に受けることが期待される伝統的な賞の付与や銅像建造において特定の不品位を排除する営為を横に置いて検討してみよう。これらと上記の「キャンセル」とが異なるのは、正しくこの責任主体性に比例的であることがわかるだろう。不買運動においては「買う」主体は自らが対価を支払ってきた、あるいは支払うだろう契約関係にあり、法的な意味で把握される私的自治の範疇に属する。翻って伝統的な賞からの不品位な人格・表現の排除や銅像となったかつての偉人の評価替えに由来する引き倒しは、まさにその「伝統」の価値(あるいは価値変動)が共同

体に属する一人一人に共有された価値（あるいは価値変動）として潜在的に分有されていることに求められよう。

⁹³ ヘイトスピーチをも取り込む社会の寛容の促進は、異なる「包括的教説」相互間の協同のための基礎構造を、国内におけるコンフリクトとして示すものと呼べようか。2015年東京法哲学研究会・法理学研究会においてこの箇所についての報告コメントをした際には本文のように明示できなかったが、本稿のベースは同報告原稿にその端緒がある。

⁹⁴ ウォルドロン『ヘイトスピーチという危害』159-162頁にあるが、「ある争点が私のアイデンティティにとって決定的に重要だということによって、私はその争点についての私の見解を政治的に交渉不可能なものとして提示する」ことによる集合的決定的挫折のことを意味する。

⁹⁵ 成原慧（2023）「キャンセルカルチャーと法」では本文に述べたキャンセルの機能をキャンセル要求が事実上果たすとしても基本的には表現の自由の枠内に属し、保護に値するとしていた。これに従えば、個人に当てられた侮蔑・名誉毀損等に該当せず、その意味で直接的でない何らかの発言が「ヘイトスピーチ」として指弾され、それに対してのキャンセル要求が機能的にキャンセル作用を果たす場合にも同様の帰結に導かれるものと思われるが、成原が比例性や手続統制を問わないそのような帰結をよしとするかは定かではない。本文では上記要求を保障した上でなお、その機能を限定する法内在的制約を論じてきた。

⁹⁶ もちろん、こうした訂正を待つことへの懸念にも、全く理由がないわけではない。害が顕在化しているケースはもとよりとして、時に一般的な意味での表現の自由の「武器化」が論じられることがあるのもその一例に挙げられよう。森口千弘（2022）「社会の分断がもたらす人権の「武器化」——マイノリティの権利の観点から」『〈分断〉と憲法 法・政治・社会から考える』は、文化闘争場面において保守的とされる側が、表現の自由の主張を盾に自らの政治主張を行い、裁判所が受け入れることをこの「武器化」の達成であるとし、これに懸念を表す。しかし、対等な道具立てで解釈を争いうる場で争っている司法的場面において、結果として特定の立場への「不都合」な帰結が生じた場合に、当該の争いの場自体の構造的問題の解消ではなく、そこで用いられる法的概念自体の使用の妥当性を論じるのは、正統性の把握という側面において転倒しているものと思われる。裁判もまた、法システム全体の中では正統性を分有する一つの機関に過ぎず、社会における「正しさ」を終局的に決する場ではない。新たな裁判にも、また裁判外にも、「正しさ」をめぐる議論のフォーラムは開かれている。したがって、この「武器化」を嘆く言説こそ、抵抗をも可能とする表現の自由を支えるエンパワメントと促進的環境形成に向けられた議論継続とを信頼することなく、「表現の自由」を我々の社会の中に位置づけることを「恐れ」ているとさえいえよう。

⁹⁷ マーサ・ヌスパウム（2010）『感情と法』（慶應義塾大学出版界会、河野哲也 監訳）355頁

⁹⁸ 同上 386頁

⁹⁹ 同上 399頁

¹⁰⁰ 表現の自由の「武器化」については前掲注96)、「歴史問題の司法化」批判については武井彩佳（2021）『歴史修正主義 - ヒトラー賛美、ホロコースト否定論から法規制まで』および同氏（2022）「歴史否定論と陰謀論」世界 961号 130-140頁を参照せよ。

¹⁰¹ 前掲注70) ミル（1971）『自由論』93-94頁

¹⁰² 同上 71頁

¹⁰³ 前掲注97) 第7章 第二節 409頁

¹⁰⁴ ジェレミー・ウォルドロン（2017）「シティズンシップと尊厳」思想1114巻。なおウォルドロンによる2008-2009 タナー講義“Dignity, Rank, and Rights”も同論点を含む。

¹⁰⁵ 同上 125-127頁を参照せよ。この二重性を、憲法学者の蟻川恒生は、たとえ文化的少数派に属する場合であろうとも、「[自らの文化的アイデンティティを公共空間の中で主張し貫徹する]ことを自ら抑制することができる個人」でありつつ、他者に対しても「抑制するよう要求することができる個人」であるとする。蟻川恒生（2016）「憲法学に「個人」像は必要か」「尊厳と身分、憲法的思惟と「日本」という問題 248頁を参照せよ。

¹⁰⁶ たとえある意味での保護に向けられた「隔離」であっても、家族集団や宗教的集団への近時の法的介入の進展を背景に、その部分社会性の限界を示すものであることは認識されているだろうし、それゆえに共同体間における移動の自由の実質的な行使可能性が鋭く問われていることはその証左である。逆もまた然りであり、「おそれ」のある表現の表現者側でさえも、刑罰を用いて物理的に（あるいはコンテンツモデレーションを介して空間的に）「隔離」することは、むしろ社会復帰を期する社会内処遇の必要性との緊張関係に立っている。OECDのメンタルヘルス関連資料における長期入院者数に見られるように、本文に記した社会内処遇の課題は刑事的な課題に限られず、「脱施設化」や共生社会一般の問題として理解される。本邦における精神疾患の長期入院者数に見られる精神疾患を抱えた者の社会からの「隔離」実践が、社会の無理解や拒絶心によって正当化されないように、社会における「おそれ」はその発言者の「隔離」を当然には正当化しないのである。最新のOECD資料情報として、下記を参照せよ。 <https://www.oecd.org/els/health-systems/mental-health.htm>（2023年7月27日最終確認）

107 マーサ・ヌスバウム「老いとスティグマと嫌悪感」(田中あや訳) 16 頁

108 前掲注 97) マーサ・ヌスバウム『感情と法』424 頁を参照せよ。ミルは社会的専制を、生活の細部に入り込むことで、逃れることを不可能にする魂の奴隷化として描き出し、考え方や慣習についての行動ルールの事実上の強制及び個性の発揮の除去の双方へと通じるものとしていた。これに対しヌスバウムは完全性の専制を、「主流を成す社会規範が、その規範と合致しない生き方に対し、法を通じて与える壊滅的な影響」として描いている。

109 Joseph Raz (1986), *Morality of Freedom*, Oxford pp396-398 を参照せよ。

110 我々は、完全ではない世界において、互いを損なうかもしれないリスクとともに、しかしそのリスクに晒された者を助け、癒し、またリスク含みかもしれない他者を許す中で、被害者と加害者の区別を乗り越えてともに生きていくより他にない時間的存在である。しかしハードな「キャンセル」は、被害と加害に至った偶然性や立場の交換可能性、また時点的な流動性を無化し、加害を与えたものとして指示された対象をその端点に還元し、公共的フォーラムの組上から取り除いてしまうだろう。

例えば、この事態は、リアル空間における書店からのいわゆる「ヘイト本」排斥要請運動と類比的である。一例として、2019 年に刊行された永江朗『私は本屋が好きでした あふれるヘイト本、つくって売らざるまでの舞台裏』を皮切りとしたヘイト本排斥要求とその後の議論を参照されたい。特定書店がヘイト本であってもその店頭で陳列する「意味」は、その意図や配置態様の巧拙、評価基準も含めて多種多様である。その「意味」あるいは機能を単純にその本を通じた差別的な雰囲気への還元することは、少なくとも、①そもその書店の陳列行動の「理解」において、また②書店という情報媒介者に求められる社会的役割を一元化する自主規制「要請」という形式において、そして③書店自身が道徳的責任主体として将来にわたる読者と社会に対する自律的な責任を果たす契機を奪う自主規制の「内容」において、各々妥当とは言えないだろう。最も重要な点として、特定の本を排斥する営為は、単に個別のヘイト本が生じさせる(かもしれない)害をパターンリスティックに除去することにどまらず、読者となるはずだった者(すなわち潜在的読者)が自らの置かれた環境に由来する関心及び能力的限界のただなかで新たな情報に開かれる場を予め奪うことで、本人の信念と真理想定を改訂する情報に触れる機会を失わせてしまうことが挙げられる。ヘイト本を手取る者にとってこそ、ヘイト本が一掃された棚ではなく、ヘイト本と反ヘイト本とが並列されている棚に触れること(あるいはそうした棚が開かれた場所における傍観者の立場としても目に付くこと)に、信念改訂上の重要な意義があることは否定し難いだろう。逆に、パターンリスティックに情報に接する本人の信念への悪影響を先回りしてリスク除去を果たすことは、本人の自律および信念改訂の機会を失わせるのであり、こうしたいわば「公共空間における情報的なゼロリスク志向」は、情報の事前の選別を放置することで、公共空間における主体への「不信」を公然と表明するものとも捉えられよう。逆説的に響くかもしれないものの、個人であれ集団であれ、リスクが特定の主体に局所化されることなくシティズンシップの尊厳を発揮できる促進的環境を構築する観点からは、リスク源となりうる候補を予め排除しようとする要求は(自己免疫的に)寛容を目指す社会像の破壊に至る。反対に、本稿で見てきたオンライン空間・コミュニケーションの特性に照らして一定の危害の限定づけを持つリスクについていえば、当該リスクが現実化する度ごとにそのリスクに晒されたものを助けつつ、その発生への寛容をまずは行使できる「タフな社会」を求めて、リスク対応の限定づけを審議し、問い直し、暫定的にでもその限界を画定させていく継続的な営みが遂行されねばならないとの帰結を導き出せるだろう。

リスクの局所化を排除しつつも、「タフな社会」を再構築する契機としてリスクをも利用する。我々は、こうした時間を通じた非難と訂正を通じた赦しの過程に開かれた存在でありうるという「信頼」を、互いに構築する努力が課されているはずである。この訂正と許しを規範的語用論から取り扱う Robert Brandom(2019), *A Spirit of Trust: A Reading of Hegel's Phenomenology* 及び Gilles Bouché (ed.)(2020), *Reading Brandom On A Spirit of Trust* の検討は次稿にて行う。



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 (2023, 近刊)「脳神経科学研究における AI・ビッグデータ取り扱いの倫理 問題となる「自律性」の種類とその対応」情報通信政策研究第 7 巻 1 号

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)

[所属] 情報学環 准教授

[所属学会] 日本法哲学会、応用哲学会、日本医事法学会、科学技術社会論学会等

Regulation of Hate Speech and Inherent Limitations (1): On the Nature of Online Space/Communications

Naoya Nagaishi*

The purpose of this paper is twofold. First, to clarify the nature of online hate speech and the characteristics of its harm from the perspective of perlocutionary and illocutionary acts that have developed in recent years; and second, to clarify the fundamental factors that determine the desirability of legal regulation.

In the online space, it is both possible and generally expected to make speech invisible and to silence speech (e.g. through so-called “content moderation”). However, by circumventing legitimate controls such as the rule of law, the quasi-regulatory action of digital platforms other than the state as a regulatory actor is problematic. These spatial characteristics limit the scope of discussions about the harms and vulnerabilities of intra- and mediated effects of speech in real space. This highlights the need for new non-regulatory and collaborative approaches to these spatial characteristics.

This paper identifies the formal characteristics of the harms of online hate speech in relation to the social and technological environment in which the speech takes place. In particular, it seeks to identify the conditions under which the harms of hate speech are irresistibly actualised.

Section 1 of this paper clarifies the scope of real-space-based arguments about the harms of online hate speech to groups in terms of (i) the relationship between collective attributes and harms to individuals, and (ii) the mechanisms of harm via the linguistic characteristics of online hate speech in non-face-to-face situations. Section 2 clarifies the premises on which the question of the punishability of online hate speech is based on the above-mentioned characteristics of harm. In particular, the distance between the mechanism of harm and the enforcement of punishment reveals the assumptions of inherent legal constraints that underlie and limit the concept of punishability. As a result, for each situation of legal regulation, the risk trade-off between harm/regulation and the risk trade-off between legal regulation and non-regulation is categorised as an inherent legal constraint. In section 3, we situate the perlocutionary and illocutionary approaches considered in this paper within the accumulation of arguments about the vulnerability of counter-speech(more speech). On this basis, at the end of the section, we present a perspective on the connection with the analysis of the philosophy of language in the next paper.

* Interfaculty Initiative in Information Studies, The University of Tokyo

Key Words : Hate Speech, Dignity, Speech Acts, Contents Moderation, the Circumstances of Politics, Identity Politics