

〔コラム〕

法の経験科学的研究から見た「民事判決情報データベース化」の意義と課題

—近年の COMPUTATIONAL LEGAL STUDIES の展開から

澤出成意人(東京大学大学院法学政治学研究科総合法政専攻・博士課程 2 年)

1 はじめに

(1) 問題の所在

2024 年 7 月 29 日、法務省の「民事判決情報データベース化検討会」(以下「検討会」という)は、「民事判決情報データベース化検討会報告書」(以下「報告書」という)を公表した。2 年弱の検討の集積であり、ニュースでも報道されるなど¹、その関心は小さくはない。全国の裁判所で出される民事裁判の判決は年間約 20 万件あるが、自動的にアクセスできるものは一部の判決に限られていた。しかし民事訴訟法改正により、裁判所は判決文を原則電磁的記録で作成しなければならなくなった(民訴法 252 条 1 項)。こうして判決文がいわば電子データ化されることを踏まえて、「検討会」は民事判決情報のデータベース化に向けて議論を重ねてきたのである。

もっとも、「検討会」は以上の民訴法改正に対応する、という消極的な目的のみにとどまっていなかった。むしろ従来になかった判決文の利活用方法の可能性を模索するために、そして民事判決情報のデータベース化のニーズを把握するために、「検討会」では有識者によるヒアリングが行われた。その中には筆者と専攻を同じくする研究者もいた。しかしヒアリングの議事録を見ると、必ずしも民事判決情報のデータベース化に対するニーズが具体的な判決文の利活用方法と関連させて提示されていないのではないかと、という印象を抱いた。

そこで本稿では、法の経験科学研究に関心を持つ者から見た、そして「検討会」のヒアリングや「報告書」では具体的に示されていない、民事判決情報データベース化そのもののニーズを明らかにしていく。アメリカ合衆国をはじめ、判決文の網羅的なデータベース化は諸外国で進行している。このデータベース化と、計算科学技術の発展が相乗効果をなして、判決情報にテキストマイニングを行って分析を行う経験科学的研究も加速度的に登場している。本稿ではこれらの判決文の計量テキスト分析の一端を紹介することを通じて、判決文の計算社会科学的分析の意義及びそれを可能にする条件を提示したいと思う。

本稿は次のように構成される。まず、2024 年 7 月 29 日に公表された「報告書」のうちとりわけ民事判決情報データ利活用の「ニーズ」がどのように把握されているかを確認すると同時に、有識者としてヒアリングを受けた研究者が提示するデータ利活用方法についても議事録を通して検討する(2)。それを踏まえて、諸外国で行われている判決文の計量テキスト分析をレビューする(3)。最後に、諸外国の研究を踏まえてデータベース化された民事判決情報の利活用方法を提示するとともに、その条件も思いつく限り示したいと思う。

(2) 本共同調査研究との関係

本論に入る前に、本共同調査研究との関係について述べておくことにする。2024 年度は本共同調査研究の最終年度であるが、本稿で扱うテーマは必ずしも主題化されているわけではない。しかし民事判決情報も、データ利活用の一場面であることは疑いようがなく、実際のデータ利活用のニーズを提示することにも意味はあろう。また、判決文は法的紛争に対して裁判所が審理した結果の報告書であり、広い意味での紛争処理と関

¹ 例えば、NHK「民事裁判判決 データベース化へ報告書まとまる 法務省の検討会」(2024 年 7 月 29 日)
<https://www3.nhk.or.jp/news/html/20240729/k10014528581000.html> 2024 年 12 月 20 日アクセス。

わる。紛争処理については、2024年9月6日に民事訴訟法の垣内秀介教授が「紛争解決の観点から見たデジタル・アーキテクチャの課題」という報告をされている。そこでは ODR (Online Dispute Resolution) など紛争解決のデジタル化について、とりわけ紛争解決システムの正当性確保の観点から、指摘をされている。後述するが、ODR については「報告書」でも指摘されており、その点で民事判決情報のデータベース化は ODR にも大きな影響を与えるものであることが予想される。

このように民事判決情報データベース化は、それなりに本共同調査研究と近接している動向である。しかし、本稿の重心はむしろ諸外国で行われている法的なテキストに対する計量テキスト分析——これらは Computational Legal Studies (以下、「ComLS」とする)を自称することがある——の紹介にある。これを紹介するのは、ComLS が理解する法とコンピュータ技術の理解(あるいはある種の「野心」)に、筆者は何か無視しえないものがあると思ったからである。2020年に Edward Elgar 社の Elgar Studies in Legal Research Methods から *Computational Legal Studies* と題する論文集が出た。その中で editor の Ryan Whalen は、現代における法とコンピュータ技術の関わり方には次の三つがあるとする²。それは第一に、サイバー法という言葉に象徴されるような、デジタル化した社会において生じる問題について法が直面するという、法的規制についての関わり方である。第二は、リーガル・テックのような法実務への技術の適用、という法実務次元での関わり方である。そして第三は、コンピュータを用いた計算科学的な研究方法が法学研究に適用される、という法学方法論の次元での関わり方である。Whalen はこの第三の関わり方を ComLS と位置付けている。Whalen のこの整理に基づくならば、ComLS が突きつけようとしている——あるいは実際に突きつけているのかもしれないが——現状は、従来の法学が前提としてきた方法論を根底から問い直すものかもしれない。ComLS を前に感じた筆者の問題意識は、デジタル社会を前にした従来の法学の問い直しあるいは反省という側面をも含んだ本共同調査研究とも少なからず共鳴するところがあるのではないかと思われるのである。

2 「検討会」における民事判決情報データ利活用の「ニーズ」把握

(1) 「検討会」の経緯及び構成メンバー

「報告書」の検討に入る前に、「検討会」について若干の紹介をしておこう³。「報告書」では、「司法制度改革審議会意見書」(2001年6月12日付)が示されている⁴。同意見書では、司法の国民的基盤の確立のためには司法に関する情報公開を推進すべきという見地に立ち、その一環として先例的価値のある判例情報をウェブサイト上で全面的に公開し提供すべきであるとされた⁵。この提言に沿って、現在裁判所ウェブサイトでは特に重要であると考えられる判決は公開されている。しかし同意見書が示された 2001 年から四半世紀近く経とうとしている現在にあっては、科学技術とりわけ AI の技術的発展は著しい。このことを踏まえて、「民事司法制度改革推進に関する関係府省庁連絡会議」(内閣官房設置)の取りまとめである「民事司法制度改革の推進」(2020年3月10日付)においては、「民事判決情報は、国民にとって、紛争発生前には行動規範となると

² Ryan Whalen, “The Emergence of Computational Legal Studies: An Introduction,” in Ryan Whalen (ed.) *Computational Legal Studies: The Promise and Challenge of Data-Driven Research* (Edward Elgar, 2020) pp.1-2.

³ 同検討会の設置経緯及び構成メンバー等については、すでに事務局を務めた法務省大臣官房司法法制部の参事官らによる紹介がなされている。石田佳世子＝大久保直輝「民事判決情報データベース化検討会報告書」の概要」NBL1277号 4-5 頁。以下の記述はこれに基づくところもある。

⁴ 「報告書」1 頁。なお「報告書」は、<https://www.moj.go.jp/content/001423117.pdf> から閲覧できる(2024年12月20日アクセス)。

⁵ 「司法制度改革審議会意見書—21世紀の日本を支える司法制度—」113 頁 https://lawcenter.ls.kagoshima-u.ac.jp/shihouseido_content/sihouseido/report/ikensyo/pdfs/iken-4.pdf 2024年12月20日アクセス。

ともに、紛争発生後には当事者による紛争解決指針の一つともなり得るものであり」、「将来的に、AI による紛争解決手続のサポートの可能性があり、その活用が国家経済の活性化にもつながり得るものであることも踏まえると、現状、先例性の高い事件や社会的に関心の高い事件等の一部の事件に限定されている民事判決情報については、今後、より広く国民に提供されるべきである」⁶という理解が示された。

このような理解に対応し、公益財団法人日弁連法務研究財団は「民事判決のオープンデータ化検討会プロジェクトチーム(PT)」(以下「PT」という)を主催した。この中で、研究者・実務家・判例雑誌出版社・判例データベース事業者による実務的な検討が行われ、提言等もなされた。「検討会」はこの流れの中で法務省が 2022 年 10 月に設置した有識者会議である。

「検討会」には、「PT」に参加していた研究者もメンバーとして入っているほか、情報法の専門家である宍戸常寿教授や巽智彦准教授、あるいは最高裁判所事務総局総務局第一課長である長田雅之氏など、法学研究者・法実務家が中心となり構成されている。また「検討会」には、山本和彦教授や町村泰貴教授など、民事訴訟の IT 化に関して研究されている民事訴訟法学者がいるほか、中原太郎教授や米村滋人教授など、民法(とりわけ不法行為法)の研究者が参加していることも注目される。実は中原・米村両教授は、「PT」の活動を踏まえ民事判決情報の適正な利活用に際して実施すべき仮名処理⁷のあり方等についてさらなる検討を加えることを目的に設置された「民事判決情報の仮名処理の在り方等に関する WG」のメンバーであった。その流れの中で「検討会」のメンバーになったのである。

(2) 「検討会」で把握されている民事判決情報データベース化の「ニーズ」

上で確認した「民事司法制度改革の推進」での民事判決情報オープンデータ化の必要性の理解が「検討会」の根幹にはあるが、「報告書」ではこれがより具体化されている。「報告書」では、民事判決情報データベース化のニーズとして次の三つを挙げる⁸。

それは第一に、司法の透明性向上・行動規範や紛争解決指針の提示、という国民基盤次元でのニーズである。これ自体は上述の「司法制度改革審議会意見書」や「民事司法制度改革の推進」と変わるところはない。要するに、憲法上判決は公開すべきであるところ、それが貫徹された形で実効されていない。現在ではウェブサイト上で重要な判決は公開されているものの、司法の透明性を向上させるという観点から見ても、それは不十分である。したがって民事判決情報を公開することによって司法の透明性を担保し、国民の法システムへの信頼を確立するとともに、国民にとって紛争解決のために必要な情報を示すことにもなる、ということである。ただし「司法制度改革審議会意見書」などにおいては判決文の公開そのものが課題であった。その意味で判決文の「オープンデータ化」のみが志向されていたと言える。しかし「検討会」はあくまで民事判決情報の「データベース化」を課題にしており、そのニュアンスは変わっているかもしれない。

第二に、一方で新規産業の創出や高度な法的サービスの提供、という(法)役務提供次元でのニーズである。データ技術の加速度的な向上に伴い、データの重要性和多様性がますます増大している。集積されたデータを用いて新たな価値を創出したり、産業や社会に新たな知見をもたらしたりする可能性があることが指摘されている。近年の法実務においても——「リーガルテック」の台頭に象徴的に現れているように——、AI による実務補助という話題はもはやその新規性を失いつつある。実務補助は、訴訟戦略のそれにとどまるのではなく、紛争解決そのものも射程に含みうる。これがいわゆる ODR の推進の目的の1つであるが、AI が紛争解決補助を十全に行うためには十分な(そして偏りのない)「学習」が必要になる。そのためには民事判決情報のデー

⁶ 「民事司法制度改革の推進について」7頁 <https://www.cas.go.jp/jp/seisaku/minjikaikaku/dai3/honbun.pdf> 2024 年 12 月 20 日アクセス。

⁷ 裁判所が判決文を裁判所ウェブサイト上で公開するにあたり、訴訟関係者の権利利益等に配慮するため、個人の氏名・住所・生年月日等を削除あるいは他の情報に置き換える処理のことを言う。

⁸ 「報告書」2-4頁。

データベース化は重要であるし、「報告書」においても「民事裁判情報を機械学習の素材として活用することにより、AI を用いて訴訟の結果を予測できるようになる時代が間もなく到来するのではないか、法律実務家の文書作成を支援するツールを開発することができるのではないかなどと様々な可能性が示唆された」⁹と記されている。

「報告書」は第三に、民事裁判情報全体の横断的分析やより精緻な統計分析を行えるようになる、という経験科学次元でのニーズも挙げている。すなわち、「先例的価値のある民事裁判情報の内容を個別に分析するとどまらず、民事裁判情報全体を通じてその傾向を分析することや同種の事案を地域ごとに分析することが可能になっており、そのためにより多くの民事裁判情報を提供する必要がある」¹⁰とされた。後に紹介するComLS の展開はまさにここで挙げられているような判決文の計量的・横断的分析を志向する。しかしこの次元でのニーズは、法実務家にとっても無視できないものである。「報告書」では、このような民事判決情報の定量的分析の成果を通じて法的役務の品質が向上し、「ひいては我が国における司法制度全体の充実・強化につながることを期待される」¹¹とされている。

(3) 社会科学者のニーズ？

如上のニーズ、とりわけ(法)役務提供次元及び経験科学次元におけるニーズは、「検討会」では有識者のヒアリングを通じて把握された。「検討会」は全部で16回の会議が開かれているが、その中で民事判決情報のニーズについては前半の第2回から第5回までの間で議題に上がっている。そして第4回会議を除く第2回・第3回・第4回会議において、有識者からのヒアリングを行っている。ちなみに第4回会議は、第3回会議までのヒアリングと議論を踏まえた、民事判決情報利活用の重要性・意義と利活用のための環境整備のあり方の課題をそれぞれ確認する、という位置付けになっている。第6回会議以降は、以上のニーズを前提として、民事判決情報の利活用に向けた制度整備のあり方に関する議論に移っていく。以下では、第5回会議までの有識者ヒアリングについて見ていくことにする。

会議	議題	発表題目	発表者	発表者の所属・専門
第2回	民事判決情報の利活用の可能性について	情報社会学の観点から	庄司昌彦	武蔵大学社会学部教授・情報社会学
		法社会学の観点から	石田京子	早稲田大学大学院法務研究科教授・法社会学
	民事判決情報の利活用の現状について	我が国における民事判決情報利活用の歴史と展望	浦野哲哉	株式会社判例タイムズ社
		民間データベース会社に対するニーズと将来展望	讃井泰雄	株式会社 LIC 代表取締役
第3回	民事判決情報の利活用の現状について	民間データベース会社の現状と将来展望	川原崎晶子	第一法規株式会社出版編集局次長兼編集第一部長
第5回	民事判決情報の利活用等について		中村裕哲	ウエストロー・ジャパン株式会社
			富所浩介	読売新聞東京本社論説副委員長

表1 「検討会」第2回から第5回までにおける有識者に関する諸情報

第2回・第3回・第4回会議における有識者のヒアリングについてまとめたものが、表1である。有識者は全部で7人いる。そのうち4人が、判例雑誌出版社や判例データベース事業者に所属している。これらの有識者は、主に法役務提供次元でのニーズについて発表している。例えば総合型法律情報システムである「判例秘

⁹ 「報告書」4頁

¹⁰ 同上。

¹¹ 同上。

書」を提供する株式会社 LIC の讃井泰雄代表取締役は、民間データベース会社に対する利用者（≡実務家）の現状のニーズを、キーワードを入れることで適切な判例が出ることとする¹⁴。その上で、民事判決情報のデータベース化の意義を、AI による判例検索をはじめとするリーガルテックの実用と関連付けて次のように述べている。

「〔民事判決文が全て公開されることになれば、〕そのオープンデータにされたものをただ提供したところでお客様〔＝法実務家〕はノイズだらけで混乱に陥るだけですから、やはり AI が検索者の意図を理解して適切に絞ってあげるといふアシストがどうしても必要になってくると思います。そのためには、判例の数が多いほど、AI のラーニングができるということですよ。今回のオープンデータ化というのは、AI の実用化を進めていく上では非常に大切な、有効なことだと思います。」¹⁵

すなわち、「使える」判例を求めている法実務家にとっては、迅速で的確な判例検索システムが必要とされる。そのため AI の精度が重要になってくる。機械学習により精度の高い AI が実装されること——つまりヘラクレス判事の生成——が重要なのであって、裁判例の網羅性（つまり判決文の「オープンデータ化」）それ自体はむしろ「ノイズだらけで混乱に陥るだけ」である。しかし AI を機械学習によりヘラクレス判事に昇華させるためには、大量の裁判例が必要になる。この点で、民事判決情報の「オープンデータ化」は「非常に大切」で「有効なこと」である。このようなニーズが示されている。

法役務提供次元でのニーズに対して、経験科学次元でのニーズについてはどのように議論されているか。有識者 7 人のうち 2 人が研究者である。1 人が情報社会学を専攻する庄司昌彦教授であり、もう 1 人が法社会学を専攻する石田京子教授である。

議事録によると、庄司教授は総務省の「自治体システム等標準化検討会」や「自治体 DX 推進検討会」の座長を歴任しており、中央省庁・地方公共団体のオープンデータ化など、行政デジタル化の推進者であるという¹⁶。庄司教授の発表は、一般的なオープンデータ化された文章を機械学習の素材として使うという利活用方法を提示することが中心である。ただしとりわけ法的に重要な示唆もしている。まず、大量の判決文を AI に読み取らせることで、実務家が文書を作成している際にそれに関連する判例あるいは法令が「おすすめ」されるというものである¹⁷。これは上述の法役務提供次元でのニーズ、というべきものである。しかしそれだけではなく、「傾向分析」、すなわち「判例を内容ごとに横断的に分析して、この人はどういうものに関わっているか、この企業はどういうところに登場しているかといったことについて分析すること」も挙げている。この分析により、「この企業は過去にこういうことがあったとか、弁護士や裁判官の傾向が分かるということもあるのだろう」という。要するに庄司教授は、判決文の「情報のタグ付け」、つまり「メタデータ」の付加がデータの価値を高めることを主張する¹⁸。

もう 1 人の研究者の有識者である石田教授も、訴訟代理人の有無などの情報をタグ付けが有意義な実証研究を産むことを主張する¹⁹。石田教授はそれだけでなく、民事判決情報がオープンになることで、司法統計の精緻化を研究者が行える、あるいは女性の司法アクセスを阻害している要因の解明にも資する可能性があ

¹⁴ 「検討会」第 2 回議事録 24 頁〔讃井氏発言〕<https://www.moj.go.jp/content/001392934.pdf> 2024 年 12 月 20 日アクセス。

¹⁵ 同上〔讃井氏発言〕。

¹⁶ 同議事録 2 頁〔事務局発言〕。

¹⁷ 同議事録 5 頁〔庄司教授発言〕。

¹⁸ 同上〔庄司教授発言〕。

¹⁹ 同議事録 12 頁〔石田教授発言〕。

るとされている²⁰。司法におけるジェンダー差を明らかにする際にも、性別によるタグ付けがあれば定量分析が容易に行えることも、後の質疑の中で示唆されている²¹。

庄司・石田両教授の発表からは、民事判決情報を単にオープンにするだけでなく、それにメタデータあるいはタグ付け(裁判官や弁護士の情報・訴訟代理人の有無・当事者の属性)があればさらにデータとしての価値が高まることが示唆されている。しかし他方、そのようなメタデータ等を付すコストは当然発生し、それゆえにそのコストの分担が問題となる。実際に「検討会」のメンバーはむしろそのことを懸念している²²。

(おそらく石田教授と同じように)法の経験的側面に関心を持つ筆者自身は、このようなメタデータが付されることそれ自体は大変有意義であると思っている。しかしそのコストについて問われれば、直ちに答えることはできない。ただ如上で示したヒアリングについて、気になる点がある。それは、経験科学次元でのニーズにおいて、民事判決情報のオープンデータ化・データベース化に内在する意義が、見過ごされているのではないかということである。むしろ、民事判決情報のメタデータが直ちに議論となり、そして具体的なコスト分担のあり方について議論の方向が移っている。石田教授が指摘したような、訴訟代理人の有無と判決結果の関係に関する定量的研究は、従来の法社会学研究においても行われてきた²³。要するにそれは、訴訟結果という被説明変数に代理人の有／無という説明変数がどれほど関連しているか、という統計分析(カイ二乗検定)により検証される課題であり、そこでは判決文を何かしらの変数として見ているわけではない。もちろん石田教授も指摘されたように、変数の間に何かしらの関連性があれば、それを説明するために「全体の判決テキストから、質的な分析をすることも必要になってくる」のだろう²⁴。しかし石田教授のこのような見方の前提にあるのは——「判決文のテキストに対する質的な検討と、量的な結果の分析」²⁵という発言に象徴的に表れているように——、「判決文のテキストを読むこと＝質的な分析／判決結果との関連を分析すること＝量的な分析」という図式なのである。この図式の中で、民事判決文の網羅的なオープン化の意義は説得的に示されていない(おそらく、示されえない)。もし民事判決文が網羅的に公開されることになったら、それを全て自分で読むことは不可能であろう。そして実際に自分で判決文を読むのだとすれば、その中の一握りの判決文を読むことに限定され、仮説を導くことになるだろう。この作業は、現代とさほど変わらなくなってしまうのではないだろうか。メタデータはもちろんあればあるほど有意義な経験科学的研究が期待されるが、その重要性が強調されても、経験科学次元におけるニーズがかえって把握されにくくなっている感が否めないのである。

石田教授は判決文を読むことを質的な分析と位置付けているが、近年の計算科学技術の発展に伴いテキストもまた計量分析の対象となっている。これがいわゆる計量テキスト分析というものである。全世界的な判決文のオープンデータ化により、法学においてもコンピューターを用いた計量テキスト分析をする研究が登場している。民事判決情報データベースの利活用の提案について、石田教授は「法社会学者でも5人集めたら5人が全く違う使い方について提言するのではないか」²⁶としていた。この指摘は正しい。以下では、石田教授が触れなかった、法学における計量テキスト分析すなわち ComLS の一端を概観することで、民事判決情報のデータベース化そのものが有する経験科学次元でのニーズを明らかにしていく。

²⁰ 同議事録 13 頁〔石田教授発言〕。

²¹ 同議事録 17 頁〔町村委員発言〕。

²² 同議事録 6 頁〔穴戸委員発言〕、18 頁〔町村委員発言〕。

²³ 例えば、齋藤宙治「民事訴訟における代理人に関する基礎的分析——10 年間の人数変化と訴訟結果の傾向——」社会科学 71 巻 2 号(2020 年)108-109 頁。

²⁴ 議事録・前掲注 13、15 頁〔石田教授発言〕。

²⁵ 同上〔石田教授発言〕。

²⁶ 同議事録 14 頁〔石田教授発言〕。

3 データとしての法

(1) 「遠読」と計量テキスト分析

諸外国の中にはすでに判決文が全て公開されているところもある。その場合、問題の設定によっては大量の判決文が分析の対象となることも考えられる。その量は、人の手では処理できないこともある。そこで、コンピュータによるテキストマイニングを通じた判決文の分析が着目されている。

ComLS ではしばしばこのようなアプローチを「遠読(distant reading)」という。ComLS の主導者である Michael A. Livermore と Daniel N. Rockmore は、次のように「遠読」を位置付ける。

遠読とは、ジャンルをはじめとする〔経験的〕事象についての大規模な次元での仮説を検証するために統計を用いて、距離をとった視点から文献を考察することである。遠読は、相対的に手作業で対処可能なテキストの詳細な分析をもって分析を進めるところの伝統的な文献学——あるいは法学——により実践されてきた、クラシックな「精読(close reading)」と対置される。²⁷

特に伝統的な法律学を例にすると、その研究は、条文や裁判例あるいは学説といった文献を詳細に読むことで進められてきた。しかしそれはあくまでも人間の手作業で行える範囲であるという意味において、一握りの(handful)テキストしか扱うことができなかった。これに対して統計的手法を用いて行う遠読は、人間の手作業では行うことが不可能な膨大なテキストを一度に扱うことができる。その点で、文献のジャンルや様式についての大規模な調査を必要とする仮説を検証することを目的とする。

なお、遠読のような発想それ自体は、新しいものではない。とりわけコンピュータの流通黎明期においても、コンピュータによる裁判過程の分析あるいは予測については議論されていた²⁸。ただそれには、テキストを、コンピュータを介して「読む」という姿勢は意識的に示されてはいない。遠読という言葉はもともと文献(史)学研究の1つの潮流²⁹から端をなすと言われているが、日本ではあまり馴染みのない言葉かもしれない。しかし遠読と同等の作業は行われてきた。それが計量テキスト分析での取り組みである。いずれにしても、自然言語処理を通じてコンピュータが解析できるようにしたテキストを、自動的に整理・分析することになる。そこでソフトウェアの利用が必須になるところ、日本では社会学者の樋口耕一教授が KH Coder というフリー・ソフトウェアを開発している³⁰。社会学において用いた計量テキスト分析が行われるようになって久しいが、法学においても用いられることはある。例えば法言語学の堀田秀吾教授は、裁判員裁判の判決文と職業裁判の判決文に現れる言葉の違いを計量テキスト分析から明らかにしている³¹。ただし民事判決文の計量テキスト分析などは、管見の限りなされていない。この背景には、「報告書」で指摘されているような判決文公開の現状があることは言うまでもない(この点については後述する)。

²⁷ Michael A. Livermore and Daniel N. Rockmore “Distant Reading the Law,” in Michael A. Livermore and Daniel N. Rockmore (eds.), *Law as Data: Computation, Text & the Future of Legal Analysis* (Santa Fe Institute, 2019) p.4.

²⁸ 例えば、Reed C. Lawlor, “What Computers Can Do: Analysis and Prediction of Judicial Decision,” *American Bar Association Journal* 49 (4) (1963) pp.337-344、宮原守男「裁判過程のコンピュータによる予測」川島武宜編集代表『(法社会学講座 5)紛争解決と法 1』(岩波書店、1972 年)363-379 頁など。

²⁹ Franco Moretti, *Distant Reading* (Verso, 2013).

³⁰ KH Coder について、樋口耕一『社会調査のための計量テキスト分析: 内容分析の継承と発展を目指して(第 2 版)』(ナカニシヤ出版、2021 年)参照。

³¹ 堀田秀吾「テキストマイニングによる判決文の分析」明治大学法学部創立 130 周年記念論文集(2011 年)472-450 頁。

(2) COMPUTATIONAL LEGAL STUDIES の展開

如上の遠読というアプローチで法テキストを「読む」ComLS の動向については、日本でもすでに飯田高教授により紹介されているところである。飯田教授によると、この 20 年前後で法学研究にも経験科学の方法による分析が広がり、その様相は empirical turn と呼ばれている。このようないわゆる Empirical Legal Studies は、法が有する何かしらの効果について関心を有している。それに対して法テキストの計量テキスト分析すなわち ComLS は、「法の構造」を解明すること、すなわち「法という存在そのものを定量的に分析しようとする」ことに主軸がある、という³²。

「定量的分析」と聞くと、観察者の視点(H・L・A・ハートがいうところの「外的視点」)から法を分析するという印象を抱くかもしれない。しかしながら、それだけに限定されない。少なくとも ComLS の提唱者たち(あるいは飯田教授)は、そのように主張している。上述の Livermore と Rockmore は、内的視点／外的視点を採用するか、そして量的アプローチ／質的アプローチを採用するかを問わず、法テキストの遠読は有益であることを主張する³³。すなわち、(1)外的視点を取り量的分析を行う場合は、大量のテキストを取り入れる必要がある。遠読を通じて、分析者が関心を有する変数に応じて「コーディング」することにより、テキストがデータの形に集約され、統計分析を容易に行えるようになる。この態度は、樋口教授の計量テキスト分析への態度と通じるところがある。(2)外的視点を取り質的分析を行う場合でも、質的分析と量的分析を融合させることで、対象をより多角的に分析できたり、あるいは旧来の(質的分析に依拠していた)問題に新たな視点を与えたりする可能性を、遠読は有している。(3)内的視点を採用する場合でも、法解釈の手助けとして遠読を用いることができることもある。例えばプライバシー侵害の四類型で著名なウィリアム・プロッサーは、何千もの判決文を読みそして消化していたことでも有名である。しかしプロッサーが生きていた時代と現代では、判決文の数は比較するまでもない。そこで、大量のテキストの言葉や言い回しの意味内容を理解することを目的として、コーパス言語学(corpus linguistics)からの道具立てを用いた試みも行われている。

このように ComLS は外的視点にとどまるものではなく、法学研究方法そのものに対する刷新を迫っているように思われる。それでは、具体的にいかなる研究がなされているのだろうか。ここで網羅的な紹介は不可能であるが、興味深いと思われるものをピックアップする形で示していきたい。

民事判決情報データベース化では、AI による訴訟結果の予測可能性も指摘されていた³⁴。ComLS のメイン・トピックの1つも、判決の予測である。例えば Charlotte S. Alexander らは、2010 年から 2017 年の間にジョージア州北部の連邦地方裁判所で提起された労働法関係の訴訟 5111 件を対象として、訴訟の結果を予測するようなモデルの開発を目指す。その前提として、訴訟の終結形態の分布や、略式訴訟(summary judgment)において裁判官が用いる法理(legal doctrine)ないしは引用する事案(case)の解明を記述的な問題として挙げる³⁵。Alexander らは、前者の問題についてドケット・シートを用いて分析している。訴訟の終結形態(ディスカバリー前／後の和解・訴え却下・原告／被告の略式判決の申立て・事実審理)の分布も、あらかじめ形態別に分類した約 1000 のドケット・エントリーをもとに、アルゴリズムがそれぞれの形態に特徴的な語を予測し、分類している。後者の問題についての素材は、略式訴訟判決文であるが、Alexander らは裁判官が依拠する法理を把握するために試行錯誤しながら、ある法理にとって特徴的な語をあらかじめ同定し、そのような語が入っているかどうかをカウントする、という方法をとった。要するに教師付

³² 飯田高「法の構造と計量分析」社会科学研究 72 巻 2 号(2021 年)4 頁。

³³ Livermore and Rockmore, *supra* note 25, p.11-14.

³⁴ 「報告書」4 頁。

³⁵ Charlotte S. Alexander et al., “Using Text Analytics to Predict Litigation Outcomes,” in Michael A. Livermore and Daniel N. Rockmore (eds.), *Law as Data: Computation, Text & the Future of Legal Analysis* (Santa Fe Institute, 2019) p.277.

き学習と教師なし学習の融合アプローチである³⁶。具体的な結果は示されていないが、その方法で判決文中に引用されている事案の頻度は示されたようである。Alexander らは最終的に訴訟結果の最も有力な予測因子として、(暫定的な結果と留保をつけて)弁護士過去の実績を挙げるが、それはあらかじめ判決文の外部にある変数などではなく、ドケットシートの内部から出てきたものだと思う。

上の Alexander らの研究でもあったように、判決文において判例や法規ないしは法理に言及することは、判決理由の論拠を示す点では当然である。つまり判決文は何かしらの論拠を参照という形で明示することで、結果の正統性を担保しようとする——このことと法的議論が反論可能性に開かれていることを是であるとする——ことは表裏一体の関係にある——文書であると言える。その意味で、引用・参照の有する意味は大きい。そこで ComLS の中には、判決文の引用・参照についての計量テキスト分析を行うものもある。例えばカナダでは、「厳格な」憲法原理としての権力分立(separation of power)が従来確立せず、カナダ最高裁判所もほとんど注意を払っていなかったところ、権利と自由憲章(Charter of Rights and Freedom)が盛り込まれる 1982 年に制定されたカナダ憲法の制定直前に、権力分立がようやく言及されるようになる。そこで「権力分立」という言葉が言及された判決文がどの程度あるのか、時代により変化しているか、関連している言葉は何か、などを計量テキスト分析により明らかにした研究がある³⁷。

同じくカナダ最高裁判所の判決文をデータとして、判例の引用ネットワークを分析したものがある³⁸。判例の引用ネットワークとは、ある判例をその後の判決が引用することで生じる関係の総体である。例えば、罪に問われた人は取り調べを合理的な時間内に限ることができる権利を有しているところ、その権利が侵害されているか否かが問題となった *R v Askov* 事件(1990 年)は、その前の判例である *Mills* 事件(1986 年)・*Rahey* 事件(1987 年)・*Conway* 事件(1989 年)をそれぞれ引用しているが、これらの判例もまた(米国の) *Wingo* 事件(1972 年)・*Ewell* 事件(1966 年)・*Strunk* 事件(1973 年)を引用している(図 1)³⁹。この研究では、1983 年から 2021 年までの約 4000 件のカナダ最高裁の判決文を題材としているが、その中では総じて 41000(憲法裁判に限れば 11115 件)に登る相互参照が見られるという⁴⁰。実際の分析では、憲法事案でランドマークになった4つの事案を素材にして、それがどのように通時的に引用されているか、そして権威スコアという指標⁴¹を用いてその意義を分析している。その結果、同じ程度引用されている判例でも、常にネットワークの中で中心を占めているものもあれば、ある時点では急激に引用されているがそのピークが過ぎると権威スコアが下がるもの、当初は引用されていないが徐々に引用されるようになっているもの、そして常にネットワークの中心にいるわけではないが通時的に一定程度引用されているもの、の四類型があることを指摘する⁴²。不法行為法など、日本でも判例法が大きな意義を有する法領域は存在する。民事判決情報データベース化が

³⁶ *Id.*, p.302.

³⁷ Vanessa MacDonnell and Keenan MacNeal, “Four Stories about Canada’s Separation of Powers Doctrine,” in Wolfgang Alschner et al. (eds.), *Decoding the Court: Legal Data Insights from the Supreme Court of Canada* (Routledge, 2024) pp.37-59.

³⁸ Wolfgang Alschner and Isabelle St-Hilaire, “Using Network Citation Analysis to Reveal Precedential Archetypes at the Supreme Court of Canada,” in Wolfgang Alschner et al. (eds.), *Decoding the Court: Legal Data Insights from the Supreme Court of Canada* (Routledge, 2024) pp.60-73.

³⁹ *Id.*, p.64-65.

⁴⁰ *Id.*, p.66.

⁴¹ この指標は、アメリカ合衆国連邦最高裁の裁判例の中で強い影響力を有する裁判例を特定し、そのような裁判例を中心としたネットワークを素描することを目的とした James H. Fowler et al. “Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court,” *Political Analysis* 15 (2007), pp.324-346 において用いられた指標である。なお、飯田・前掲注 30、14-15 頁も参照。

⁴² Alschner and St-Hilaire, *supra* note 36, p.68-73.

実施されれば、例えば判例のネットワークを可視化することもできるようになるとされるし、判例が下級審において(どのように)参照されているかなども分析することが可能になろう。

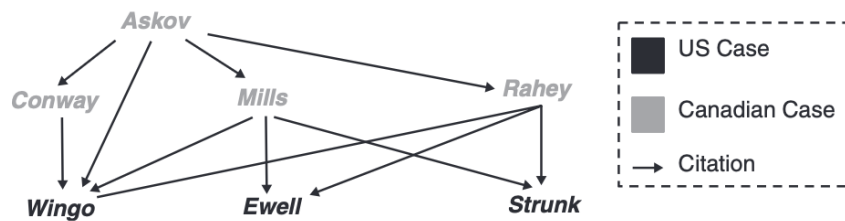


図1 判例の引用ネットワーク⁴³

上に示した判例の引用分析などは、従来の法学研究とも親和的な側面がある。この点で、法の計量テキスト分析は、実際には従来の法学研究に取って代わるものというよりは、その補完的存在にとどまっているというのが妥当な見方であるといえよう。しかしこれを逆に言えば、伝統的な法学研究が用いてこなかった経験科学的手法が、たとえ補完的なものであっても無視できない程度の存在感を有するようになったのだとも言える。ここにも法学研究の empirical turn が顔を覗いている、というわけである。このことは、上で紹介してきた諸研究が行われてきたコモン・ローの法文化に対してのみならず、大陸法圏に対しても言えるかもしれない。制定法国であり日本法にも大きな影響を与えてきたドイツにおいても、法学における計量的アプローチないしはAI・機械学習の利用は議論されるようになってきている⁴⁴。ドイツでも判決文の電子化が進み、遠読による判決文分析も可能になりつつある。例えば、1951年から2017年までに公式判例集に掲載された3308件のドイツ連邦憲法裁判所の判決文を、確率的に関連する語をクラスタリングするトピックモデルにより分析したものがある⁴⁵。このトピックモデルは、人間が分類するようなトピックとは異なる分類を示すことが多い。むしろコンピュータによって作り出されるクラスタリングの中には、人間が理解できないものがある。それを踏まえて著者らはトピックモデルの結果を過信するべきでなく、関心のあるトピックを定義するための布石として捉えた方が良いという見方を示す⁴⁶。その上で、ドイツ連邦憲法裁判所が用意する裁判手続である憲法異議(Verfassungsbeschwerde)と具体的規範統制(konkrete Normenkontrolle)の間で異なるトピック(法領域)が持ち込まれているか、という課題を設定し、これを検証している。具体的には、トピックモデルにより自動的に生成されたトピックを、専門家の助言を経て個々の法領域(「民法」「社会法」など)に分類・ラベリングする。こうして法領域ごとにトピックが分類されたので、憲法異議と具体的規範統制の判決文の間の法領域の分布を量的に捉える、という方法をとっている。その結果、いくつかの法領域については両制度の間で有意に出現頻度が異なることがわかった。すなわち、租税法・社会法・公務員法は有意に具体的規範統制において登場する法領域である。一方、表現の自由・データ保護法は有意に憲法異議において登場する法領域であることがわかった⁴⁷。このような経験的知見を得る基礎としてトピックモデルが用いられているのであり、その前提には判決情報がデータベースとして公開されていることが不可欠であったわけである。

⁴³ 図1の出典は、*Id.*, p.65.

⁴⁴ *Vgl.*, Tobias Gump und Marc Pierre Schneider, „Methoden der Künstlichen Intelligenz in der Rechtswissenschaft,“ *Zeitschrift für Digitalisierung und Recht* (2021), S.155–174.

⁴⁵ Luisa Wendel et al. “From Modeled Topics to Areas of Law: A Comparative Analysis of Types of Proceeding in the German Federal Constitutional Court,” *German Law Journal* 23 (2022), pp.493–531.

⁴⁶ *Id.*, p.504. 著者らはこのような態度を示すが、それは決して消極的なものではない。むしろ大量の判決文——それらを全て精読することは不可能である！——から何らかの整理をしようとする際に、トピックモデルは有益であるとしている。

⁴⁷ *Id.*, p.523. なお、刑法はどちらの法制度でも登場するが、徐々に憲法異議において登場してきているとする。私法もどちらの法制度でも登場しているが、それはラベリングに問題があった(私法に特徴的な語を反映できていなかった)とする。

判決文の文体(writing style)は、例えばメールを書くといった日常生活上の書き言葉とは異なるものである。しかしその文体は永久不変あるいは普遍的なものでもなく、時代や地域によって異なることもまた事実である。日本の裁判所の判決文でも「畢竟独自の見解に過ぎない」という定型文句はよく見られたものである。ComLS がコーパス言語学からの影響も受けていることに鑑みても、この文体の問題は1つの主題である。上述のように判決文の文体が法文化・法伝統から影響を受けているとなると、異なる法文化・法伝統が接触するトランスナショナルな裁判所はいかなる文体なのか、という疑問が浮かぶ。ComLS の主導者の1人でもある Jens Frankenreiter は欧州司法裁判所(ECJ)の判決文を題材として、この問題に取り組む⁴⁸。具体的な指標として Frankenreiter は一文の長さや語彙の多様性、あるいは音節の数や機能語(functional words)の分布などを用いながら、1952年から2015年までのECJの全ての意見を対象とした文体の分析を行う。Frankenreiterによれば、従来ECJの判事はフランス人が多かったことから、フランス的な法文化がECJの判決文に影響を与えていた。しかしEUの拡大に伴い他の法文化を有する国から判事が任官されるようになり、現代では必ずしもかつてのようなフランスの法文化の影響はなくなっているという議論がなされている、という⁴⁹。そこで如上の指標を用いた通時的な文体の変遷を、判事の退任・任官時期と関連させながら分析している。その結果、ECJの判決文は時期により文体が異なるが、EUの拡大に伴って文体が変わるというパターンが見られるわけではなく、そして依然としてフランス語圏からの判事がそれ以外の判事と比べてECJの文体の発展に影響を与えていることを示している⁵⁰。日本でも民事判決文のデータベース化が進めば、判決文の文体についての裁判所間の差異や時系列分析を網羅的に行うことができるようになるだろう。判決文内の構造も、判決文をいかなる形式で公表するかによっては明らかにできる可能性はある。

近年の社会科学ではテキストの中に含まれる感情的なトーンをコンピュータが自動的に識別・分析する感情分析などもあるが、これと似たような形でテキスト内に含まれるイデオロギー性や党派性を何らかの指標を用いて測定する分析も行われている。とりわけ ComLS の主戦場であるアメリカでは、リアリズム法学の舞台でもあることから想像されるように、裁判官の党派性・イデオロギーと訴訟結果の関係については関心が高い。こうした関心から、アメリカの連邦地方裁判所及び連邦控訴裁判所における裁判官の政治的党派性・イデオロギーと判決文の相関関係を定量的に調べたものがある⁵¹。これは、経済学者が書いた論文のイデオロギーを測定した指標を判決文に流用して分析したものである。経済学者の書いたものにおいてはイデオロギー性が浮かび上がったのに対して、判決文からは裁判官の党派性を検出することはできなかった。この結果の理由について著者は、二つの可能性を挙げている⁵²。一つは、現状の言語モデルあるいはテキスト分析の技術では検出不可能であったという可能性である。もう一つは、法律論におけるいかなる党派的影響も(少なくとも法の素人にとっては)知覚することが困難であるという可能性である。これは、ある問題に公的な意味をもたせる機能を有する法言語(諸概念・思想・執筆作法)は、裁判官の間では党派性を超えて相対的に均一(homogenous)であることによる、とする。もちろん第一の可能性もあるだろうが、著者自身は第二の可能性の検討に重心を置いている。というのも著者は、上述のように、経済学者の論文という明示的には政治性を醸し出さないテキストの論述パターンから明確にリベラル／保守のいずれのイデオロギー集団に属するかを予測できることを実証した研究を参照し、その方法を判決文に応用したからである。このようなイデオロギー測定分析は翻って法

⁴⁸ Jens Frankenreiter, "Writing Style and Legal Traditions," in Michael A. Livermore and Daniel N. Rockmore (eds.), *Law as Data: Computation, Text & the Future of Legal Analysis* (Santa Fe Institute, 2019) pp.153-189.

⁴⁹ *Id.*, pp.157-159.

⁵⁰ *Id.*, p.188.

⁵¹ Marion Dumas, "Detecting Ideology in Judicial Language," in Michael A. Livermore and Daniel N. Rockmore (eds.), *Law as Data: Computation, Text & the Future of Legal Analysis* (Santa Fe Institute, 2019) pp.383-405.

⁵² *Id.*, pp.400-403.

言語の有する特質を解明することができるのではないか、という期待を抱かせるような研究であることは間違いない。

この他にも様々な判決文の計量テキスト分析があり、それをレビューすることそのものが一苦勞になると思われる。しかし如上に紹介した諸研究を見るだけでも、民事判決情報データベース化の経験科学次元でのポテンシャル及びそれを発掘するニーズがあることがわかるだろう。必ずしもメタデータが付されていなくとも、コーディング次第では、大量の判決文を類型化可能な場合もある。同様に、コンピュータに読み込ませることで予期しなかったパターンを発見できることもある。要するに従来の法社会学・法律学の枠組みではあまり焦点が当たらなかった問題を発見できるポテンシャルを、民事判決情報データベース化そのものが有しているのである。

4 おわりに

筆者は不法行為法に興味を有しており、例えば伝統的通説とされる相関関係説がどの程度裁判実務に影響を有しているか、などは関心のある問題である。このような経験的課題を、計量テキスト分析を用いて扱おうとしたこともある。しかし現状、そのような分析が依拠するデータは、判例データベース事業者を通じて入手できる判決文に限られている。時代ごとに公開されている判決文の数も異なるので、飯田教授がいうところの「法の通時的構造」⁵³を素描することすら不可能である。そこに網羅性がないことは承知の通りであり、それこそが「検討会」の問題意識であった。

飯田教授もこの点を捉えて、分析対象の代表性の確保に注意を促す。「判決のデータベースの場合、収録されアクセス可能なのは、実際に出された莫大な量の判決のごく一部である。時代によって収録範囲が異なることもある」⁵⁴。しかしより問題は、データの偏りがあることそれ自体というよりも、データの偏りを分析者が把握・説明できないことなのではないか。判例データベース事業者がデータベースに裁判例を掲載する基準が明示されていればまだ偏りを説明でき、その点で留保を明確にできるのだが、それすらないのが現状である。民事判決情報のデータベース化が実現すると、この点の問題をクリアできるのはかなり大きい。上で紹介した判例の言及分析やネットワーク分析なども、より充実したものができるようになることは間違いない。

このような判決文の計量テキスト分析が法学の方法全体の刷新——ComLS が陰に陽に見せるような——を促すものなのか、あるいは逆に、法律学が伝統的な方法に固執し ComLS は単なるブームに留まるのか、それは不明である。筆者の予想はそれらの中庸である。すなわち、従来の法学の方法も維持しながらも、吸収できる方法——計量テキスト分析は、情報縮減の側面も有する——が取り入れられる、という穏当な未来が来るのではないかと考えている⁵⁵。

この穏当な未来の到来のためにも、民事判決情報データベース化にあたりいくつかの条件が必要である。すでに報告書では、判決データベースの低廉な価格での提供については指摘されている⁵⁶。また、当事者の性別など、判決文からは直ちには分別不能なメタデータを付与することについては、いかなるメタデータを付与すべきか、さらにメタデータ付与のコスト分配はいかにすべきか、についても「検討会」で少々議論に登っていた。「報告書」ではメタデータについては確定的なことは示されておらず、今後の検討が求められる。

これ以外にも少なくとも二つ、判決文の計量テキスト分析をするにあたっての課題がある。第一に、判決等のテキストを構造化されたデータとして入手可能なものにするということである。例えばデジタル庁が提供する法

⁵³ 同上、4頁。

⁵⁴ 飯田・前掲注 30、21 頁。

⁵⁵ 同旨、成原慧「民事判決情報データベース化・オープンデータ化の意義と課題——司法における AI の活用に向けて」法の支配 215 号(2024 年)105 頁。

⁵⁶ 「報告書」32 頁。

令検索・閲覧システム「e-Gov 法令検索」(<https://laws.e-gov.go.jp>)では、XML 形式で法令のダウンロードが可能である。XML(eXtensible Markup Language)とは、文書構造を記述する言語の1つであり、データの論理構造・属性・機能がタグによって指定される。例えば上述の「e-Gov 法令検索」から民法を XML 形式でダウンロードする。民法 399 条は、XML 形式では図 2 のように記述される。<Part Num="3">が第3編を、<Chapter Num="1">が第1章を、<Section Title>が第1節を、<Article Num="399">が 399 条をそれぞれ示すことにより、民法 399 条が民法第3編第1章第1節にあることを構造的に示している。このような XML 形式のテキストデータは、法の計量テキスト分析には有用である。例えば、飯田教授は民事法令を XML 形式でダウンロードし、そこから法令名・条文番号・条文のテキストデータを抽出している。このデータを元に、民事法の「共時的構造」をプロット的に描く⁵⁷。法令はこのように XML 形式で取得可能であるのに対して、判決文はここまで構造化された形での入手はできない状況にある。とりわけ判決理由は、ある程度内容的まとまりに区切られてナンバリングされた形で構成されている。その意味で構造的な文書であるが、現状その構造を自動的に記述できない。民事判決文が原則電子化されるならば、二次分析にも利用しやすい XML 形式での公開も検討に値しよう。

```
<Part Num="3">
  <PartTitle>第三編 債権</PartTitle>
  <Chapter Num="1">
    <ChapterTitle>第一章 総則</ChapterTitle>
    <Section Num="1">
      <SectionTitle>第一節 債権の目的</SectionTitle>
      <Article Num="399">
        <ArticleCaption>(債権の目的)</ArticleCaption>
        <ArticleTitle>第三百九十九条</ArticleTitle>
        <Paragraph Num="1">
          <ParagraphNum/>
          <ParagraphSentence>
            <Sentence Num="1" WritingMode="vertical">債権は、金銭に見積もることができないものであっても、その目的とすることができる。</Sentence>
          </ParagraphSentence>
        </Paragraph>
      </Article>
    </Section>
  </Chapter>
</Part>
```

図 2 民法 399 条の XML 形式による記述

判決文の入手に関連して、第二に、判決文が一括入手できることも重要である。Lexis Nexis では複数のファイルをまとめてダウンロードできるサービスがあるが、日本の判例データベースではそのようなサービスはない。そこで Ruby や Python 等のプログラミング言語を用いたデータを自動収集するプログラム(スクレイピング)により判決文をダウンロードすることが考えられるところ、そもそも大量データのダウンロードを利用契約違反としている判例データベース事業者も多い⁵⁸。しかし判決文の計量テキスト分析は、大量の判決文を手でできてこそ意義を発揮する。また、ComLS では判決文の入手はスクレイピングを用いて行われていることも多い。大量の判決文の入手を容易にすることも、求められよう。

以上本稿では、民事判決情報データベース化の意義を、ComLS の研究成果の一端を紹介しながら示し、さらにデータベースを用いて実際に分析するにあたって必要な条件——判決データの構造化・大量入手可能性の確保——を提示した。「検討会」の委員は法学者が多いが、研究者も民事判決情報データベース化に対してのステークホルダーに他ならない⁵⁹。ステークホルダーとしての社会学者による一意見として思っていただけでも幸いである。

⁵⁷ 飯田高「法の構造の定量的分析」日本法社会学会編『法社会学の最前線』(有斐閣、2023 年)9頁。

⁵⁸ 森田果『実証分析入門——データから「因果関係」を読み解く作法』(日本評論社、2014 年)287 頁参照。

⁵⁹ 法律学的な関心からも、判決文がごく一部しか公開されていない現状は問題視されている。例えば上野達弘「研究者から見た裁判実務～知財司法がリードする未来～」コピライト 758 号(2024 年)2-39 頁。