

LISTENING FROM AFAR: AN ALGORITHMIC ANALYSIS OF TESTIMONIES FROM THE INTERNATIONAL CRIMINAL COURTS

Renana Keydar†

Abstract

Despite the recognized importance of witness testimony in addressing systematic violence and human rights violations, reflected in the participation of large numbers of witnesses in international legal processes, establishing facts based on oral testimonies in international criminal tribunals remains a contentious matter.

The article develops a new model for assessing judicial attention to and engagement with testimonial narratives, in particular of victims of sexual violence, by conceptualizing the testimonies as “textual datasets.” This article presents the results of an algorithm-based approach for analyzing testimonial corpora, applying a generative statistical model known as unsupervised topic modeling. I employ LDA topic modeling for empirically assessing the international courts’ capacity to “listen” to large quantities of witness testimonies. Harnessing the large number of testimonies in international criminal trials, I use topic modeling in order to explore latent themes and semantic fields that could benefit the legal process and its critical scholarly appreciation.

This article proposes Automated Content Analysis, in particular topic modeling method, as a novel method to assist scholars and practitioners in making sense of complex legal cases, involving large amounts of testimonies, documents, and data, while preserving the voice and vocabulary of the individual witness.

This article highlights the potential of topic modeling methods, rooted in Natural Language Processing and Digital Humanities, to overcome critical impediments in empirical legal studies. It demonstrates the method’s capacity to transform both as a practical heuristic mechanism that can be employed during the legal proceeding, and in its ex-post analysis in legal scholarship.

† Assistant Professor of Law and Digital Humanities, The Hebrew University of Jerusalem. The author wishes to thank the participants of the Conference on Data Science and Law, ETH Zurich, June 2019 and anonymous reviewer for their insightful comments. The author also thanks Meital Zalberg for her instrumental assistance in computational analysis.

TABLE OF CONTENTS

I.	Introduction	56
II.	Automated Content Analysis for Empirical Legal Research	60
	A. Background	60
	B. Topic Modeling	61
	C. Topic Modeling in Legal Scholarship – Pros and Cons	64
	D. Topic Modeling the Quantitative Turn – Listening to the Plurality of Voices.....	66
III.	Methodology	67
	A. Corpus	67
	B. Pre-Processing	69
	C. Number of Topics.....	69
IV.	Topic Modeling a Trial – Analysis of the Model.....	70
	A. Naming the Topics	70
	B. Overview of Topics	71
	C. Topic Distribution Over the Course of the Trial.....	72
	D. Semantic Fields of Rape.....	74
	E. Assessing the Gaps between Victims’ Testimony and Judicial Opinion.....	78
V.	Conclusion	82

I. INTRODUCTION

In June 2018, the Appeals Chamber of the International Criminal Court (ICC) decided by a narrow majority of three to two to acquit Congolese politician Jean-Pierre Bemba Gombo from charges of war crimes and crimes against humanity, for which he was unanimously convicted two years earlier.¹ The “seismic decision,”² which overturned an important victory for sexual violence victims, offered a *de novo* review of the facts of the case.³ This unusual appellate intervention was particularly surprising given that the conviction had been reached after a four-and-a-half year-long trial in which seventy-seven witnesses gave testimony and 773 items of evidence were introduced.⁴ The trial transcript was thousands of pages long, and the judgement spanned 364 pages.⁵

1. Diane Marie Amann, *In Bemba and Beyond, Crimes Adjudged to Commit Themselves*, EJIL: TALK! (June 13, 2018), <https://www.ejiltalk.org/in-bemba-and-beyond-crimes-adjudged-to-commit-themselves/>.

2. *Id.*

3. *Id.*; Leila Nadya Sadat, *Fiddling While Rome Burns? The Appeals Chamber’s Curious Decision in Prosecutor v. Jean-Pierre Bemba Gombo*, EJIL: TALK! (June 12, 2018), <https://www.ejiltalk.org/fiddling-while-rome-burns-the-appeals-chambers-curious-decision-in-prosecutor-v-jean-pierre-bemba-gombo/>; Jennifer Trahan, *Bemba Acquittal Rests on Erroneous Application of Appellate Review Standard*, OPINIO JURIS (2018), <http://opiniojuris.org/2018/06/25/bemba-acquittal-rests-on-erroneous-application-of-appellate-review-standard/> (last visited Nov. 25, 2018).

4. Trahan, *supra* note 3.

5. Alex Whiting, *Appeals Judges Turn the ICC on its Head with Bemba Decision*, JUST SECURITY (June 14, 2018), <https://www.justsecurity.org/57760/appeals-judges-turn-icc-head-bemba-decision/> (explaining that the trial opened on 22 November 2010, closing oral statements were heard on 12–13 November 2014 and the

This volume of evidence is not uncommon in the international criminal courts.⁶

In this sense, the Bemba case is emblematic of what I elsewhere identified and labeled as the *quantitative turn* in ICL,⁷ referring primarily to the participation of a large number of witnesses and victims in legal proceedings dealing with mass atrocity.⁸ However, the Bemba appeal decision is also interpreted to be a watershed moment for the failure of quantity in the courtroom.⁹ The frustrating gap between the long, complicated and costly litigation—which included testimonies of dozens of fact witnesses, experts, and victims of systematic rape—and the judicial resolution of the case requires a model that is able to deal with such multi-dimensional data. Recent technological developments in the field of Artificial Intelligence (AI) and Law, Natural Language Processing (NLP) and Automated Content Analysis (ACA) can assist in bridging this gap between quantity and judicial decision-making.

Instead of searching for ways to reduce the quantity of testimonies, this article aims at understanding the impact of quantity on the legal process. More concretely, this article asks how the quantity of data, and specifically the quantity of *viva-voce* testimonies, affects the justice process. The question it seeks to answer is whether we can examine if an over-exposure to suffering narratives (or more broadly, to large amounts of data) could result in judicial inattentiveness. This is not an inconsequential concern. Given the reliance of international criminal tribunals on eyewitness testimony for evidentiary purposes,¹⁰ inattentiveness in the face of multiple acts of human narration could impede the fact-finding process. It could also compromise the commitment to meaningful participation of victims.

In order to answer these questions and concerns this article develops a novel model of “distant listening” for empirically assessing the international courts’ capacity to “listen” to large quantities of witness testimonies. The model harnesses the large number of testimonies, which it considers as datasets, analyzing them using a novel ACA method called *topic modeling*¹¹ in order to explore latent themes and semantic fields that could benefit the legal process and its critical scholarly appreciation.

judgment was rendered eighteen months later on 21 June 2016); *Case Information Sheet - The Prosecutor v. Jean-Pierre Bemba Gombo*, ICC, <https://www.icc-cpi.int/CaseInformationSheets/BembaEng.pdf>.

6. See generally Renana Keydar, *Mass Atrocity, Mass Testimony, and the Quantitative Turn in International Law*, 53 L. SOC. REV. 554 (2019) (discussing the increasingly large numbers of witnesses in international criminal proceedings).

7. *Id.*

8. *Id.*

9. See Sadat, *supra* note 3, (criticizing the decision of the Appeals Chamber).

10. Nancy A. Combs, *Deconstructing the Epistemic Challenges to Mass Atrocity Prosecutions*, 75 WASH. & LEE L. REV. 223 (2018); Nancy A. Combs, *Grave Crimes and Weak Evidence: A Fact-Finding Evolution in International Criminal Law*, 58 HARV. INT’L L. J. 47 (2016); NANCY A. COMBS, *FACT-FINDING WITHOUT FACTS: THE UNCERTAIN EVIDENTIARY FOUNDATIONS OF INTERNATIONAL CRIMINAL CONVICTIONS* (2010); Megan A. Fairlie, *The Abiding Problem of Witness Statements in International Criminal Trials*, 50 N.Y.U. J. INT’L L. & POL. 75 (2017); John D. Ciorciari & Anne Heindel, *Victim Testimony in International and Hybrid Criminal Courts: Narrative Opportunities, Challenges, and Fair Trial Demands*, 56 VA. J. INT’L LAW 265 (2016); Keydar, *supra* note 6.

11. See generally Yuening Hu et al., *Interactive Topic Modeling*, 95 MACH. LEARN. 423, 423–69 (2014) (providing an overview of topic modeling).

Given the legal and public controversy around the Bemba trial, the article thematically focuses on *a more established and less disputed* case-study from the International Criminal Tribunal for the Former Yugoslavia (ICTY), *Prosecutor v. Dragoljub Kunarac et al.* (Case No. It-96-23-T and IT-96-23/1-T), also known as the “Foča rape camps” trial. Like the Bemba case, the Kunarac trial is also emblematic of the quantitative turn in ICL: through dozens of testimonies, many by rape victims, the *Kunarac* case addresses the systematic enslavement and rape of young Muslim women (as young as twelve years old) by Serb forces during the ‘ethnic cleansing’ campaign in the municipality of Foča between the years 1992-1993.¹² The Kunarac trial is noteworthy for being the first case solely to focus on the rape, torture and mistreatment of women during the armed conflict in Yugoslavia.¹³ It is also the first case to successfully prosecute rape as a crime against humanity, and it is the first case to consider the phenomena of ‘mass rape’ in the context of Bosnia-Herzegovina.¹⁴

This article is not, doctrinally speaking, about the rape in Bosnia or rape as an international crime.¹⁵ It is a consideration of how courts and other justice mechanisms respond to the experience of quantity in the courtroom. Underlying this study is a concern with the role of the human voice in the justice making process, and the ways in which the large-scale, mass violence, such as the mass rapes in Bosnia-Herzegovina, are represented, understood, and processed in ‘official’ knowledge-producing sites.¹⁶

Employing topic modeling, a method originating in the field of NLP and ACA, the article will analyze the entire corpus of the Kunarac trial transcripts, while maintaining a linguistic focus on the individual narrative unit of the testimony, preserving the “voice” of the victims. Extracting the latent topics in the trial, the model will enable a comparison between the themes emerging from the trial transcripts, the narratives of the witnesses and victims, and the themes that comprise the judicial decision that adjudicated them, thus aiding in exposing potential gaps between the two.

My hypothesis is that this newly developed model of “distant listening” will: 1) expand our understanding of episodes of mass human rights abuse, in particular sexual violence, by enlarging the data we have at hand; 2) improve the work of international courts that address human rights violations, by providing

12. See Christopher Scott Maravilla, *Rape as a War Crime: The Implications of the International Criminal Tribunal for the Former Yugoslavia’s Decision in Prosecutor v. Kunarac, Kovac, & Vukovic on International Humanitarian Law*, 13 FLA. J. INT’L L. 321, 321–42 (2000) (discussing the Foča rape camps, where it is estimated that more than 20,000 women were raped during the Balkan War).

13. *Id.* at 322.

14. Doris Buss, *Prosecuting Mass Rape: Prosecutor v. Dragoljub Kunarac, Radomir Kovac and Zoran Vukovic*, 10 FEM. LEG. STUD. 91, 91–99 (2002).

15. Vast scholarship deals with rape and sexual violence as an international crime. For socio-legal analysis of the Kunarac case, see, for example: Buss, *supra* note 14; Kirsten Campbell, *The Trauma of Justice: Sexual Violence, Crimes Against Humanity and the International Criminal Tribunal for the Former Yugoslavia*, 13 SOC. LEG. STUD. 329 (2004); Kirsten Campbell, *Rape as a ‘Crime Against Humanity’: Trauma, Law and Justice in the ICTY*, 2 J. HUM. RIGHTS 507 (2003); S. McClellan Dorothy & Knez Nikola, *Prosecuting Rape, Sexual Enslavement, and Genocide in Time of War: Southeastern Europe 1991–1995*, 5 INT’L J. SOC. SCI. 79 (2016); and, more generally, Janet Halley, *Rape in Berlin: Reconsidering the Criminalisation of Rape in the International Law of Armed Conflict*, 9 MELB. J. INT’L L. 78 (2008).

16. Buss, *supra* note 14.

an empirical assessment of their fact-finding mission; and 3) contribute to the capacity of courts to listen to testimonial narratives dealing with human rights abuses.

The strength of the method demonstrated in the article is its applicability to the broader legal field. The topic modeling method, which we introduce in this article, can assist scholars and practitioners in making sense of complex legal cases, involving large amounts of testimonies, documents and data.¹⁷ Whether these are complex litigation cases, both criminal and civil,¹⁸ mass torts cases, cases before the ad-hoc international mass claims commission and tribunals,¹⁹ or class actions under the Alien Tort Statute and related U.S. legislation.²⁰

Given the relative novelty of the topic modeling approach, particularly in its application to the legal field, Section II will provide a background discussion of topic models, describing some of the essential features and potentialities of this statistical method, and its appeal to empirical legal studies. Section III will describe the methodology, the corpus and the training of the model. Section IV will then analyze the results of the model, namely the topics. The analysis will 1) demonstrate how the LDA algorithm provides a reliable view of a trial through topic modeling of its transcripts, and 2) show the potential topic modeling to illuminate gaps in themes emerging from the legal process, reflected in the court transcripts—in particular, victim testimony—and the judicial opinion expressed in the judgment. Taken together, the article will demonstrate the potential contribution of topic modeling to the field of empirical legal studies, both as a practical heuristic mechanism that can be employed during the legal proceeding,²¹ and in its ex-post analysis in legal scholarship.

17. Stuart Ford, *Complexity and Efficiency at International Criminal Courts*, 29 EMORY INT'L L. REV. 1 (2015).

18. *Complex Litigation*, NATIONAL CENTER FOR STATE COURTS, <https://www.ncsc.org/Topics/Civil/Complex-Litigation/Resource-Guide.aspx> (last visited Nov. 18, 2018). Complex Litigation is the category of cases requiring more intensive judicial management. Complexity may be determined by multiple parties, multiple attorneys, geographically dispersed plaintiffs and defendants, numerous expert witnesses, complex subject matter, complicated testimony regarding causation, procedural complexity and more. While civil cases such as mass torts and class actions are the classical examples of complex litigation, many of the cases before the international criminal tribunals exhibit this procedural complexity that requires special judicial attentiveness.

19. Lea Brilmayer, *Understanding IMCCs: Compensation and Closure in the Formation and Function of Intentional Mass Claims Commissions*, 43 YALE J. INT'L L. 273, 273–314 (2018).

20. Natalie R. Davidson, *Alien Tort Statute Litigation and Transitional Justice: Bringing the Marcos Case Back to the Philippines*, 11 INT'L J. TRANSITIONAL JUSTICE 257 (2017); Natalie R. Davidson, *Shifting the Lens on Alien Tort Statute Litigation: Narrating US Hegemony in Filártiga and Marcos*, 28 EUR. J. INT'L L. 147 (2017).

21. For an overviews of the use of some technological means currently used to organize the enormous amount of factual information generated in cases before the international courts, see, for example, David Pimentel, *Technology in a War Crimes Tribunal: Recent Experience at the ICTY*, 12 WILLIAM MARY BILL RIGHTS J. 715 (2004); Yvonne McDermott, *Inferential Reasoning and Proof in International Criminal Trials*, 13 J. INT. CRIM. JUSTICE 507 (2015).

II. AUTOMATED CONTENT ANALYSIS FOR EMPIRICAL LEGAL RESEARCH

A. Background

The field of ACA is in its infancy but has the potential to revolutionize both the practice and the study of the law.²² While more and more legal databases are available online,²³ we simply do not have the human power to read and study them to their fullest. However, recent advances in NLP, computational linguistics and computer science have laid the ground for systemic automated textual analysis.²⁴

“To this end, Machine Learning (ML) researchers have developed probabilistic topic modeling, a suite of algorithms that aim to discover and annotate large archives of documents with thematic information.²⁵ Topic modeling algorithms are statistical methods that analyze the words of the original texts to discover the themes that run through them, how those themes are connected to each other, and how they change over time.”²⁶

This research develops a new empirical strategy to address the quantitative turn in international law and to empirically assess the courts’ capacity to listen to the multitude of testimonies. The study employs a statistical process known as *unsupervised topic modeling*²⁷ to a dataset containing the entire corpus of transcripts from the international criminal trial dealing with the conflict-related crimes of sexual violence. Understanding large collections of unstructured text remains a persistent problem in all data-related sciences, including the law. Unlike qualitative analysis based on information retrieval, where researchers know what they are looking for, topic models are attractive because they offer a formalism for exposing a corpus’ themes by discovering groups of words that often appear together in documents (the namesake “topics”).²⁸

Employing methods of ACA, such as topic modeling, allows expansion beyond the human and readerly scale of a “close reading” methodology limited to a small number of texts, and complement it with “distant reading”²⁹ that analyzes large corpora of text.

Crucially, unlike other empirical methods, topic models do not require the translation or reduction of documents into binary or numerical form,³⁰ making them especially attractive to the analysis of complex narratives, such as

22. For an overview on the historical development of content analysis in legal scholarship and practice, its epistemology and methodologies, see: Mark A. Hall & Ronald F. Wright, *Systematic Content Analysis of Judicial Opinions*, 96 CAL. L. REV. 63 (2008); Chad M. Oldfather, Joseph P. Bockhorst & Brian P. Dimmer, *Triangulating Judicial Responsiveness: Automated Content Analysis, Judicial Opinions, and the Methodology of Legal Scholarship*, 64 FLA. L. REV. 1189 (2012).

23. Hall & Wright, *supra* note 22, at 70.

24. Oldfather et al., *supra* note 22, at 1189.

25. David M. Blei, *Probabilistic Topic Models*, 55 COMM. ACM 77 (2012).

26. *Id.*

27. *Id.*

28. Yuening Hu et al., *Interactive Topic Modeling*, 95 MACH. LEARNING 423, 424 (2014).

29. Franco Moretti, *Conjectures on World Literature*, 1 NEW LEFT REV. 54, 56 (2000).

30. See, e.g., David S. Law, *Constitutional Archetypes*, 95 TEX. L. REV. 153, 164 (2016) (discussing the difficulty of converting the free-form nature of constitutional preambles into numerical data).

testimonies, as well as other free-form legal texts. Employing topic modeling technique to the testimonies provides a unique way to explore the content of these testimonies as a dataset, providing both the close reading as well as the distant, corpus-wide, reading.

B. Topic Modeling

Topic modeling is an exploratory technique, useful for imposing order upon large bodies of textual data and for discovering information that helps analysts see beyond their priors. Topic modeling algorithms are a suite of ML methods for discovering hidden thematic structure in large collections of documents.³¹ With a collection of documents as input, a topic model can produce a set of interpretable “topics” (i.e., groups of words that are associated under a single theme) and assess the strength with which each document exhibits those topics.³² Topic models enable researchers to code text collections that are too large to code by hand—a topic model will estimate a coding instrument and situate each document within it.³³

Furthermore, a topic model might uncover topics that a researcher using hand coding methods might not otherwise have seen.³⁴ Using NLP and ML algorithms that detect semantic structure patterns, large bodies of text units can be classified into semantically similar clusters *without* human direction, requiring only that the researcher will later assign a label to the clusters based on the key words and the documents the analytics identify as the core of a semantic cluster.³⁵

This article employs Latent Dirichlet Allocation (LDA), which was the first topic modeling algorithm and today is considered the standard algorithm.³⁶ LDA is a statistical model of language. It assumes that there are a set of topics in a collection (the number is specified in advance), where a topic is formally defined as a distribution over a vocabulary.³⁷ Terms that are prominent within a topic are those that tend to occur in documents together more frequently than one would expect by chance. In LDA, each document exhibits those topics with different proportions.³⁸

The model described below contains topics about “consent” (topic 0), “military attacks” (topic 4) and “evidence” (topic 8). By a topic being “about” a subject, “we mean that those distributions over the vocabulary place high

31. David M. Blei, *Probabilistic Topic Models*, 55 COMM’N OF THE ACM 77, 82 (2012).

32. *Id.* at 78.

33. *Id.* at 83.

34. Paul DiMaggio, Manish Nag & David Blei, *Exploiting Affinities Between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding*, 41 POETICS 570, 576 (2013).

35. J.B. Ruhl, John Nay & Jonathan M. Gilligan, *Topic Modeling the President: Conventional and Computational Methods*, 86 GEO. WASH. L. REV. 1243, 1248–49 (2018).

36. David M. Blei, Andrew Y. Ng & Michael I. Jordan, *Latent Dirichlet*, 3 J. MACH. LEARN. RES. 993, 996 (2003); Blei, *supra* note 31 at 77–84; *see also* Jonathan Chang et al., *Reading Tea Leaves: How Humans Interpret Topic Models*, ADVANCES IN NEURAL INFO. PROCESSING SYSTEMS 288–296 (2009) (noting that an LDA algorithm is implemented in Python Gensim package, employing a wrapper of MALLET).

37. Blei et al., *supra* note 36, at 996.

38. *Id.* at 1011.

probability on words that an analyst would interpret as related to the subject.”³⁹ Transcripts that discuss the Accused version, for example, will exhibit the topics 2 and 4, “military attacks” and “military command” respectively; transcripts that discuss military expert testimony will exhibit topic 4 “military command” and transcripts of rape victims will exhibit topic 9 “factual framework.” We emphasize that these topics are not known in advance. *Figure 1* illustrates the thirty top ranked terms from topics uncovered in our collection of documents from the Kunarac trial.

Figure 1: Table of Topics

	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
1	Kovac	fws	day	Muslim	unit	case	injury	judge	witness	house
2	apartment	Kunarac	man	Foca	command	crime	case	interpretation	accused	remember
3	time	rape	time	Serb	military	article	person	witness	defense	time
4	Vukovic	witness	village	building	group	law	medical	honor	evidence	soldier
5	girl	soldier	Kunarac	part	order	prosecutor	examination	give	Kovac	room
6	remember	house	Foca	work	army	act	rape	person	fact	man
7	food	accused	road	town	soldier	person	doctor	understand	rape	happen
8	witness	Vukovic	area	area	detachment	international	report	read	Kunarac	woman
9	flat	woman	Rogoj	kp-dom	Kunarac	judgement	cyst	time	indictment	told
10	soldier	girl	night	party	report	criminal	finding	evidence	case	day
11	told	Foca	august	weapon	force	sentence	expert	Vukovic	time	Partizan
12	asked	man	command	village	responsibility	rape	opinion	defense	testimony	girl
13	cafe	case	vehicle	side	Foca	convention	consequence	remember	claim	school
14	good	Kova	July	April	rank	torture	fact	today	part	rape
15	place	Partizan	told	ethnic	herzegovina	victim	dr	problem	point	Kunarac
16	year	July	person	photograph	defense	statute	conclusion	cross	Vukovic	night
17	Foca	February	hour	municipality	territorial	offense	victim	paragraph	prosecutor	child
18	lived	testified	house	rally	independent	violation	change	correct	Muslim	witness
19	town	dp	morning	military	Bosnia	circumstance	experience	moment	July	give
20	unit	school	mentioned	family	man	committed	woman	marked	event	describe
21	building	testimony	assignment	show	officer	common	intercourse	explain	relation	brought
22	Kostic	August	happen	Serbian	time	art	professor	put	circumstance	daughter
23	mentioned	indictment	kalinovik	member	tactical	Yugoslavia	trauma	asked	action	Serb
24	ensure	evidence	arrived	uniform	area	humanity	point	good	victim	village
25	Muslim	Kovac	line	house	part	appeal	find	investigator	issue	Foca
26	leave	paragraph	battalion	mosque	task	relevant	year	day	put	asked
27	occasion	Osmana	knew	hear	jna	conflict	traumatic	detail	argument	bus
28	knew	incident	body	map	combat	force	examined	yesterday	Foca	long
29	corresponds	apartment	gaga	conflict	leader	state	psychological	date	count	occasion
30	blank	stated	2nd	neighborhood	correct	considered	stress	examination	position	husband

39. DiMaggio et al., *supra* note 34, at 578.

for LDA are derived by taking a Bayesian probabilistic perspective, encoding the topics and the per-document topic proportions as hidden random variables in a hierarchical probabilistic model and then approximating the conditional distribution of those variables given an observed collection of documents.⁴⁰ In this article, we analyze the output of such algorithms. Given the texts, LDA inference produces a set of topics (figure 1), and for each document, an estimate of its topic proportions and to which topic each observed word is assigned.⁴¹

C. *Topic Modeling in Legal Scholarship – Pros and Cons*

Applying statistical techniques directly to the text of court transcripts dovetails with the growing availability of legal materials in machine-readable form and the increasing sophistication of computational text analysis.⁴² Digitalization shifts “the bottleneck of content analysis from the costs of access and tedious human coding to the need for good theory, sound methodology, and software.”⁴³

A common strategy in content analysis is to produce a set of themes (based on research questions, theoretical priors, or perusal of a subset of texts), create a coding sheet, and then code texts by reading them (or, more often, by having research assistants read them).⁴⁴ The limitations of this approach are (a) that it is impractical when corpora are very large; (b) the more analytically sophisticated are the research questions, the harder it is to achieve acceptable levels of intercoder reliability; and (c) the approach presumes that the researcher knows what is worth finding in the texts before having analyzed them. Another strategy involves using computer programs to search texts for keywords (selected based on research questions or theoretical priors) and comparing subsets of texts with respect to the prevalence of those keywords.⁴⁵ This approach requires the researcher to circumscribe the scope of exploration a priori and, by treating each instance of a term as equivalent, it also assumes that there is a single, inherent, fixed meaning to a term, which is neither contextual nor relational.⁴⁶

At least two perennial challenges arise from the use of ACA—first, “projects face a tradeoff between large-scale inquiry focused on ‘thin’ observations (e.g., voting agreement, participation, coalition size, length of legal documents), and smaller-scale studies that involve the coding of more abstract and nuanced concepts.”⁴⁷ “This inverse relationship between breadth and depth

40. Blei et al., *supra* note 36, at 997.

41. *Id.* at 1012.

42. See Michael A. Livermore, Allen B. Riddell & Daniel N. Rockmore, *The Supreme Court and the Judicial Genre*, 59 ARIZ. L. REV. 837, 862–63 (2017) (discussing the increasing use of computational text-analysis within the context of legal documents).

43. KLAUS KRIPPENDORFF, *CONTENT ANALYSIS: AN INTRODUCTION TO ITS METHODOLOGY* 50 (4th ed. 2004).

44. DiMaggio et al., *supra* note 34, at 577.

45. *Id.*

46. *Id.*

47. Michael Evans et al., *Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research*, 4 J. EMPIRICAL LEGAL STUD. 1007, 1008 (2007).

limits researchers' ability to understand the dynamics of the judicial system more fully."⁴⁸ Second, content analysis almost always raises questions about coding reliability, and this is especially true in judicial research.⁴⁹

Topic Modeling methods overcome these critical impediments. In essence, they flip the research process, using the "machine" to classify and trace topics in the text corpus first and only then relying on the human to interpret the results.⁵⁰ These methods classify enormous amounts of data into semantically similar clusters *without* human direction,⁵¹ and, crucially, without reducing the documents into binary or numerical form.⁵² This makes topic modeling especially attractive to the analysis of complex narratives and free-form texts, such as court transcripts, and testimonies in particular.

Following its initial introduction in the emerging field of Digital Humanities,⁵³ topic modeling is increasingly applicable to complex literary texts, including novels,⁵⁴ plays,⁵⁵ and poetry.⁵⁶ Using this computational method to supplement traditional literary interpretation draws on the method's capacity to analyze context- and genre-dependent linguistic and semantic structures.⁵⁷ It attests to the method's appeal to other text-based disciplines like the law.

Despite making inroads into the social sciences and the humanities, and despite its apparent appeal, topic modeling remains relatively new to legal scholarship.⁵⁸ Recent years have seen a modest rise in the application of ACA techniques to legal corpora.⁵⁹ But, the scholarly, as well as practical, potential

48. *Id.* at 1039.

49. *Id.*

50. Ruhl et al., *supra* note 35.

51. *Id.* at 1248.

52. *See id.* (describing how the process works by detecting semantic structure patterns and failing to include any reference to reducing documents into binary or numerical form).

53. *See* Matthew L. Jockers & David Mimno, *Significant Themes in 19th-Century Literature*, 41 *POETICS* 750, 751 (2013) (describing how humanities scholars have begun using topic modeling).

54. *Id.* at 752.

55. *See generally* Christof Schöch, *Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama*, 11 *DIG. HUMAN. Q.* 1 (2017) (discussing French plays in the context of topic modeling).

56. *See generally* Borja Navarro-Colorado, *On Poetic Topic Modeling: Extracting Themes and Motifs From a Corpus of Spanish Poetry*, 5 *FRONT. DIG. HUMAN.* 1, 3–5 (2018) (discussing poetry in the context of topic modeling).

57. *See generally id.* (discussing computational methods through poem analysis).

58. A search in Westlaw's Law Reviews and Journals database of keywords "Automated Content Analysis" and "Topic Modeling" in Westlaw database returned 12 relevant results of articles employing ACA techniques to legal corpora. All 12 articles are from the last decade 2008-2018, with most results (7) from 2016-2018. Some preliminary examples for the application of automated content analysis and topic models include: Keith Carlson, Michael A. Livermore & Daniel Rockmore, *A Quantitative Analysis of Writing Style on the U.S. Supreme Court*, 93 *WASH. L. REV.* 1461 (2015); David S. Law, *Constitutional Archetypes*, 95 *TEX. L. REV.* 153 (2016); David S. Law, *The Global Language of Human Rights: A Computational Linguistic Analysis*, 12 *LAW ETHICS HUM. RIGHTS* 111 (2018); Michael Livermore, Allen Riddell & Daniel Rockmore, *The Supreme Court and the Judicial Genre*, *ARIZ. L. REV.* 837 (2017); Chad M. Oldfather, Joseph P. Bockhorst & Brian P. Dimmer, *Triangulating Judicial Responsiveness: Automated Content Analysis, Judicial Opinions, and the Methodology of Legal Scholarship*, 64 *FLA. L. REV.* 1189 (2012); Ruhl, Nay, & Gilligan, *supra* note 35; Michael A. Livermore, A. Riddell & Daniel Rockmore, *Agenda Formation and the U.S. Supreme Court: A Topic Model Approach*, *SSRN* (2016); Daniel Taylor Young, *How Do You Measure a Constitutional Moment: Using Algorithmic Topic Modeling to Evaluate Bruce Ackerman's Theory of Constitutional Change*, 122 *YALE L.J.* 1990 (2012).

59. *Id.*

of the method is still far from being tapped.⁶⁰ Scholars began employing topic modeling to legal texts in attempt to map out huge amounts of data and to create a “bird’s eye view” of patterns across corpora: Supreme Court judgments,⁶¹ appellate court decisions,⁶² presidential directives,⁶³ and constitutions⁶⁴ are some examples, as well as stylistic analysis of judicial opinions.⁶⁵ Importantly, topic models are not only about organizing and categorizing.⁶⁶ More deeply, topic models can also help lawyers and legal scholars conceptualize law and legal institutions.⁶⁷

Taking the topic modeling methodology one step further, we employ the method in order to get not a bird’s eye view, but rather a close-up analysis that will provide a deeper, more nuanced understanding of the most intimate pieces of text in the legal field—witness testimonies—and their relation to the judicial decision.

We emphasize that the complexity of language implies that ACA methods do not replace careful and close reading of texts. Rather, the proposed method is best thought of as amplifying and augmenting careful reading and thoughtful analysis.⁶⁸ Instead of thinking of the computer as replacing the human, the ideal is to combine machine-enabled ‘distant reading’ with more traditional interpretative techniques of close reading.⁶⁹

D. *Topic Modeling the Quantitative Turn – Listening to the Plurality of Voices*

The main motivation for introducing topic modeling to the legal domain is to help us “make sense of it all!”⁷⁰ “Topic models are a means of organizing large bodies of knowledge into coherent structures that help us navigate the corpus of information.”⁷¹ This is particularly needed when dealing with complex litigation, such as international criminal trials.⁷²

The multitude of topics, perspectives and information which is a source of anxiety in any complex trial is intrinsic to the subject matter of international criminal litigation that is characterized by large numbers of perpetrators, victims, types of offences, related events, multiple locations, and long duration.⁷³ The complexity of international trials is on the substantive level of the crimes and the multifaceted factual framework necessary to prove them; but it is also

60. *Id.*

61. Livermore et al., *supra* note 42, at 837.

62. Oldfather et al., *supra* note 22.

63. Ruhl et al., *supra* note 35.

64. Law, *supra* note 58.

65. Livermore, et al., *supra* note 42; Carlson et al., *supra* note 58.

66. Ruhl et al., *supra* note 35.

67. *Id.*

68. Justin Grimmer & Brandon M. Stewart, *Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts*, 21 *POLIT. ANAL.* 267 (2013).

69. FRANCO MORETTI, *DISTANT READING* (2013).

70. Ruhl et al., *supra* note 35, at 1250.

71. *Id.*

72. *Id.* at 1240–41.

73. Keydar, *supra* note 6, at 10.

logistical, with international trials encompassing voluminous bodies of investigative material that the parties and the judges must process.⁷⁴

Moreover, a virtue of topic modeling is its deep affinity to the central insight in literary analysis of texts and that is that texts do not necessarily reflect a singular perspective but are often characterized by polyphony, the co-presence of competing voices, perspectives or styles of expression, within a single text, whether it is a legal document, a literary novel or any other cultural means of expression.⁷⁵ A particular strength of topic modeling is its ability to capture polysemy and disambiguate different uses of a term, based on the context (other terms) in which it appears.⁷⁶ In their emphasis on relationality, topic models capture the insight, shared by linguistics and literary scholars, that meanings emerge out of relations rather than residing within words. Thus, many terms may appear in more than one topic within a given corpus of documents.⁷⁷ This is particularly useful when dealing with international trials featuring large number of testimonies, presenting diverging and even contrasting accounts, narratives and semantic meanings, as will be discussed in chapter 4.

III. METHODOLOGY

A. *Corpus*

For the purpose of this article, we focus on the ICTY's *Kunarac* case, which addresses the atrocities committed in the municipality of Foča during the 1992-1995 war in Bosnia and Herzegovina between the Serb and Muslim forces.⁷⁸ During and after the takeover of the municipality by Bosnian-Serb forces in April 1992, Muslim and other non-Serb inhabitants were subjected to a widespread and systematic pattern of abuses, designed to remove the majority of them from the municipality.⁷⁹ Muslim girls and women were particularly targeted during this campaign and became victims of repeated rapes and atrocious sexual abuses.⁸⁰

On February 22, 2001, Trial Chamber II of the ICTY delivered its decision in the *Kunarac* case, convicting the three male defendants of the rape, torture and enslavement of Muslim women and girls in the municipality of Foča:

What the evidence shows is that the rapes were used by members of the Bosnian Serb armed forces as an instrument of terror, an instrument they were given free rein to apply whenever and against whomsoever they wished. What the evidence shows is that it was

74. Pimentel, *supra* note 21, at 717–20.

75. DiMaggio et al., *supra* note 34, at 591.

76. See, Roberto Navigli, *Word Sense Disambiguation: A Survey*, 41 ACM COMPUT. SURV. CSUR 10, 3–4 (2009) (stating that, in computational linguistics, word-sense disambiguation refers to the process by which we identify which sense of a word is used in a sentence, when the word has multiple meanings).

77. DiMaggio et al., *supra* note 34, at 587.

78. Maravilla, *supra* note 12, at 323–25.

79. *Id.*

80. Buss, *supra* note 14, at 91–99; Bridging the Gap - Foča, Bosnia and Herzegovina, UNITED NATIONS: INTERNATIONAL CRIMINAL TRIBUNAL FOR THE FORMER YUGOSLAVIA, <https://www.icty.org/en/outreach/bridging-the-gap-with-local-communities/foc> (last visited Nov. 27, 2018).

possible for the Serb forces to set up and maintain a detention centre for scores of Muslim women . . . from which women and young girls were taken away on a regular basis to other locations to be raped. . . . What the evidence shows are Muslims, women and girls, mothers and daughters together, robbed of their last vestiges of human dignity, women and girls treated like chattels, pieces of property at the arbitrary disposal of the Serb occupation forces, and more specifically, at the beck and call of the three accused.⁸¹

Many of the cases tried at the ICTY are exceedingly complex and difficult.⁸² The Kunarac case is no exception. The Kunarac trial lasted 58 days, over a period of eight months.⁸³ The trial featured 33 prosecution witnesses, 29 defense witnesses (including the accused Kunarac; the two co-accused, Kovac and Vukovic, did not testify), and one witness called by the chamber.⁸⁴ The prosecution presented 132 exhibits and the defense presented 130.⁸⁵

The trial is characterized then by a plurality of participants, many of whom are victims of heinous sexual violence as part of a systematic campaign of terror and ethnic persecution.⁸⁶ This makes a paradigmatic case of the quantitative turn, foregrounding questions of quantity.⁸⁷ Particularly, the *Kunarac* case raises questions of quantity in the context of sexual violence through the engagement with narratives of victims.

Current practices for conflict and post-conflict sexual violence investigations focus almost exclusively on the introduction of testimonial evidence in international courts.⁸⁸ This makes testimony an indispensable component of the judicial justice making process.⁸⁹ It is thus crucial for the analysis of the *Kunarac* case to employ a method that will maintain the focus on the sensitive, complex, and contextual linguistic traits of the act of testimony without reducing it to numbers and codes.

81. Prosecutor v. Dragoljub Kunarac Radomir Kovac and Zoran Vukovic, Case No. IT-96-23-T & IT-96-23/1-T, Decision for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law, 6559–60 (Int'l Crim. Trib. for the Former Yugoslavia Feb. 22, 2001).

82. Pimentel, *supra* note 21, at 720.

83. The trial opened on 20 March 2000 with closing arguments on 20–22 November 2000. Judgment was handed down on 22 February 2001. The accused appealed against the Trial Chamber judgment and sentence. On 12 June 2002 the Appeals Chamber rendered its judgment affirming the judgment and sentence handed down by the Trial Chamber.

84. *Case Information Sheet: The Prosecutor v. Kunarac, Kovac, and Vukovic*, INT'L CRIM. CT., <https://www.icc-cpi.int/CaseInformationSheets/BembaEng.pdf> (last visited Jan. 28, 2020).

85. *Id.*

86. *Id.* at 5.

87. See Keydar, *supra* note 6, at 2 (describing the quantitative nature where there is a “large number of witnesses and victims.”).

88. Marie-Helen Maras & Michelle D. Miranda, *Overlooking Forensic Evidence? A Review of the 2014 International Protocol on the Documentation and Investigation of Sexual Violence in Conflict*, 2 GLOB. SECUR. HEALTH SCI. AND POL'Y 10, 12 (2017).

89. See *id.* (explaining the weight the testimonial evidence holds in cases).

B. Pre-Processing

All the transcripts from the *Kunarac* trial are openly available on the ICTY website.⁹⁰ The transcripts are organized by date and do not include any reference to their content.⁹¹ Consequently, despite the wealth of information on the website, it is very poorly accessible for query purposes.⁹²

As a first step, we downloaded all the transcripts and identified the testimonial components within each transcript. While in general we aimed for the entire raw data, we attempted to avoid procedural “noise” that would have biased the model. So, for example, we did include substantial party statements such as opening and closing statements as well as significant exchanges between the parties and the Trial Chamber. Given the focus of the article on the capacity of courts to “listen” to the multitude of voices in the legal process, we did not include written materials that were submitted during the trial, such as motions and requests, only the oral component of the sessions. The only exception is the judgment, which was included in the model for the purpose of comparing its topical content to the topics that were extracted from the transcripts.

Beyond pre-processing the data, we tagged the transcripts according to their metadata, categorizing each transcript to either prosecution or defense. Each testimony was then tagged as expert witness, fact witness, victim or accused, and the gender identity of the witness (male/female).

We used some conventional techniques to improve the quality of the data we collected, including tokenizing and lemmatization. In the last round of processing, we filtered out “stopwords,” very common words like articles, conjunctions, or forms of the verb “to be” and words that are highly frequent across the corpus.⁹³ At the end of the processing phases, our corpus included 211 files (0–210), containing 310,374 significant words, 11,300 of which were tokens.

C. Number of Topics

Topic models can produce any number of topics that the researcher specifies. One chooses the number based on interpretability and analytic utility. Selecting the number of topics is one of the most problematic modeling choices in finite topic modeling.⁹⁴ Not only is there no clear method for choosing the number of topics in human-interpretable terms, but the degree to which LDA is

90. U.N International Residual Mechanism for Criminal Tribunals, CRIM. TRIBUNAL FOR THE FORMER YUGOSLAVIA, *Kunarac et al.*, <http://www.icty.org/case/kunarac/4#trans> (last visited Jan. 24, 2020) (most public transcripts include redacted parts, which is considered a “soft” protective measure).

91. *Id.*

92. *Id.*

93. See Alexandra Schofield, Mans Magnusson & David Mimno, *Pulling Out the Stops: Rethinking Stopword Removal for Topic Models*, 2 PROC. OF THE 15TH CONF. OF THE EUROPEAN CHAPTER OF THE ASS'N FOR COMPUTATIONAL LINGUISTICS 432, 433 (2017) (arguing for minimizing the removal of stopwords from the corpus in the pre-processing phase).

94. Hanna M. Wallach, David M. Mimno & Andrew McCallum, *Rethinking LDA: Why Priors Matter*, 22 ADVANCES IN NEURAL INFO. PROCESSING SYSTEMS 1973, 1980 (2009).

robust to a poor setting of the number of topics is not well-understood.⁹⁵ Although nonparametric models provide an alternative, they lose the substantial computational efficiency advantages of finite models.⁹⁶

Think of the model as a lens for viewing a corpus of documents. The point is to identify the lens through which one can see the data most clearly. Just as different lenses may be more appropriate for long-distance or middle-range vision, different models may be more appropriate depending on the analyst's substantive focus. When topic modeling is used to identify themes and assist in interpretation, there is no statistical test for the optimal number of topics or for the quality of a solution. Thus, the test of the model as a whole is its ability to identify a number of substantively meaningful and analytically useful topics, not its success in optimizing across all topics.⁹⁷

We tested the algorithm's parameters for a variety of options stretching from as little as 7 topics and up to 50 topics. We narrowed this down to a range of 7–15 topics according to topic coherency and human interpretability. Too many topics produced “thin” topics without cohering around a core idea and too little topics resulted in very broad general “basket topics” that did not provide useful information. After centering on the 7–15 range of topics we tested several options, and ultimately settled on a 10–topic model, which proved the most computationally stable, since it produced the same topics again and again across repeated iterations.

IV. TOPIC MODELING A TRIAL – ANALYSIS OF THE MODEL

A. Naming the Topics

After determining the number of topics, the LDA algorithm generated the model, producing a list of words from the corpus, sorted into the clusters or “topics” (as demonstrated in *figure 1* and visualized in *figure 2* above).⁹⁸

Importantly, given the LDA method's unsupervised nature, the topics generated by the computer are not named. They are a result of a statistical—not an interpretative—process. The algorithm “reads” the corpus of transcripts, and based on computations of hundreds of texts, it generates a new text of the ten most important themes dealt with in the trial, represented in the word clouds in *figure 2* above. Now, it is the task of the scholar to interpret this “new” text based on her expert domain knowledge through a process of naming the topics. This was done by assigning a label to each topic based on the key words and the documents, which we identified as the core of a semantic cluster. In general,

95. *Id.*

96. *Id.*

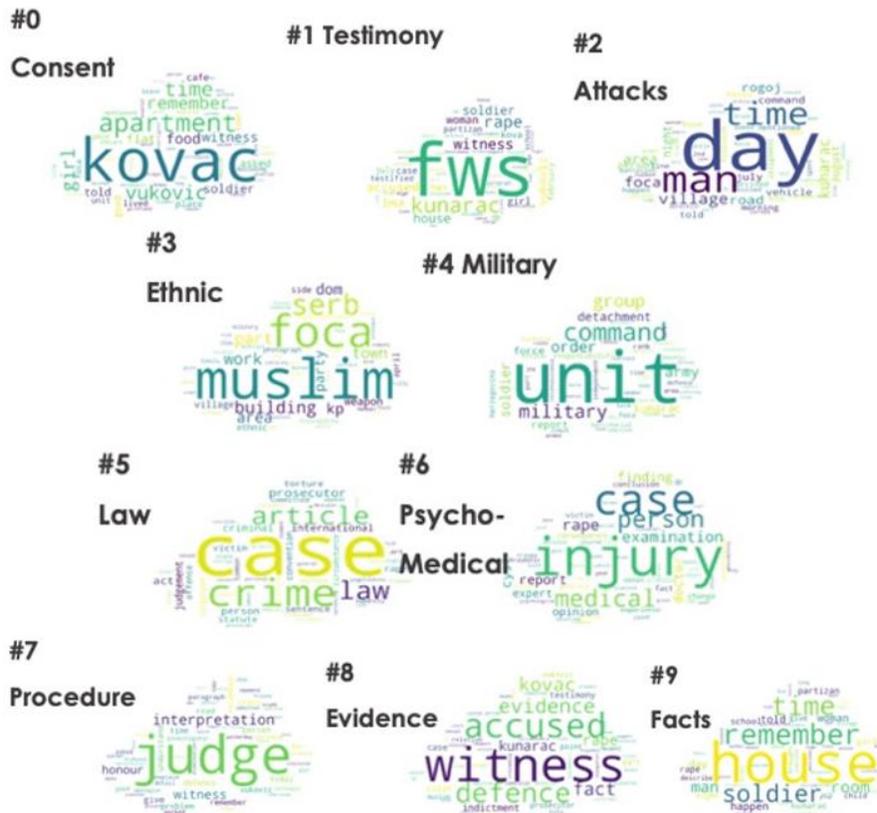
97. DiMaggio et al., *supra* note 34, at 582–83.

98. The LDA result can be interpreted as a distribution over topics. For example, in the *Kumarac* case study it would look like this:

0=0.08855770459279548, 1=0.07697199648000046, 2=0.125428335996302, 3=0.10009352690459682, 4=0.08927440923009342, 5=0.07403124645595478, 6=0.050702636123636395, 7=0.1365367195384379, 8=0.08527529066554118, 9=0.17312813401264157. This result suggests that topic 9 “factual framework” has the strongest representation in this corpus: 17% of the corpus.

where possible, the term chosen to “describe the topic” is one of the highest ranked terms within the topic. Like any clustering technique, the topic model method should be employed as a heuristic tool in combination with additional information by a research team that includes subject-area experts.⁹⁹ For that reason we validated our naming decision by consulting with four scholars familiar with the subject matter. *Figure 4* shows the results of the interpretative naming process based on the computational analysis of the corpus.

Figure 4: Naming the Topics



B. Overview of Topics

Figure 4 provides an overview of the themes comprising the Kunarac trial: 0-consent (or ‘soldier-girl relations’—see discussion below); 1-testimony (nouns

99. Justin Grimmer & Brandon M. Stewart, *Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts*, 21 POL. ANALYSIS 267, 285 (2013) (describing topic models in LDA); *Id.*

and verbs relating to the act of giving testimony or describing the testifiers¹⁰⁰); 2-attacks on civilian targets; 3-ethnic tensions; 4-military command structure; 5-legal framework; 6-psycho-medical; 7-court procedure; 8-evidence; 9-factual framework.

A taxonomy of topics provides some understanding of the types of topics that a model of trial transcripts can extract. The model discovers themes corresponding to two layers of information: “meta-trial” topics that describe core elements of a trial: legal framework, factual framework, testimony and evidence, and trial procedure. Other topics form a different category of “case-specific vocabulary” and are more reflective of the concrete themes of the Kunarac case. These include, for example, the topic of attacks on civilian population and military command and structure that pertain to the establishment of the contextual element of armed conflict in the charges of crimes against humanity.¹⁰¹ Another example is the topic of ethnic tensions, which reflects the discussion of the ‘ethnic cleansing’ campaign during the Balkan War.

A closer examination of the terms in each topic produces yet a more nuanced taxonomy. Topics 0, 1, 2, 3, 8 and 9 are case-specific, with terms relating to the concrete facts of the case, such as names of people, geographical locations and dates. Topics 5 and 7 include procedural and legal terms that have broad applicability to trials in general. Topics 4 and 6, dealing with military command structure and the psycho-medical respectively, are somewhat ambiguous in the sense that their themes are rooted in the details of the Kunarac case, and yet the terms of each topic can be descriptive of sexual violence in armed conflicts more generally.

C. Topic Distribution Over the Course of the Trial

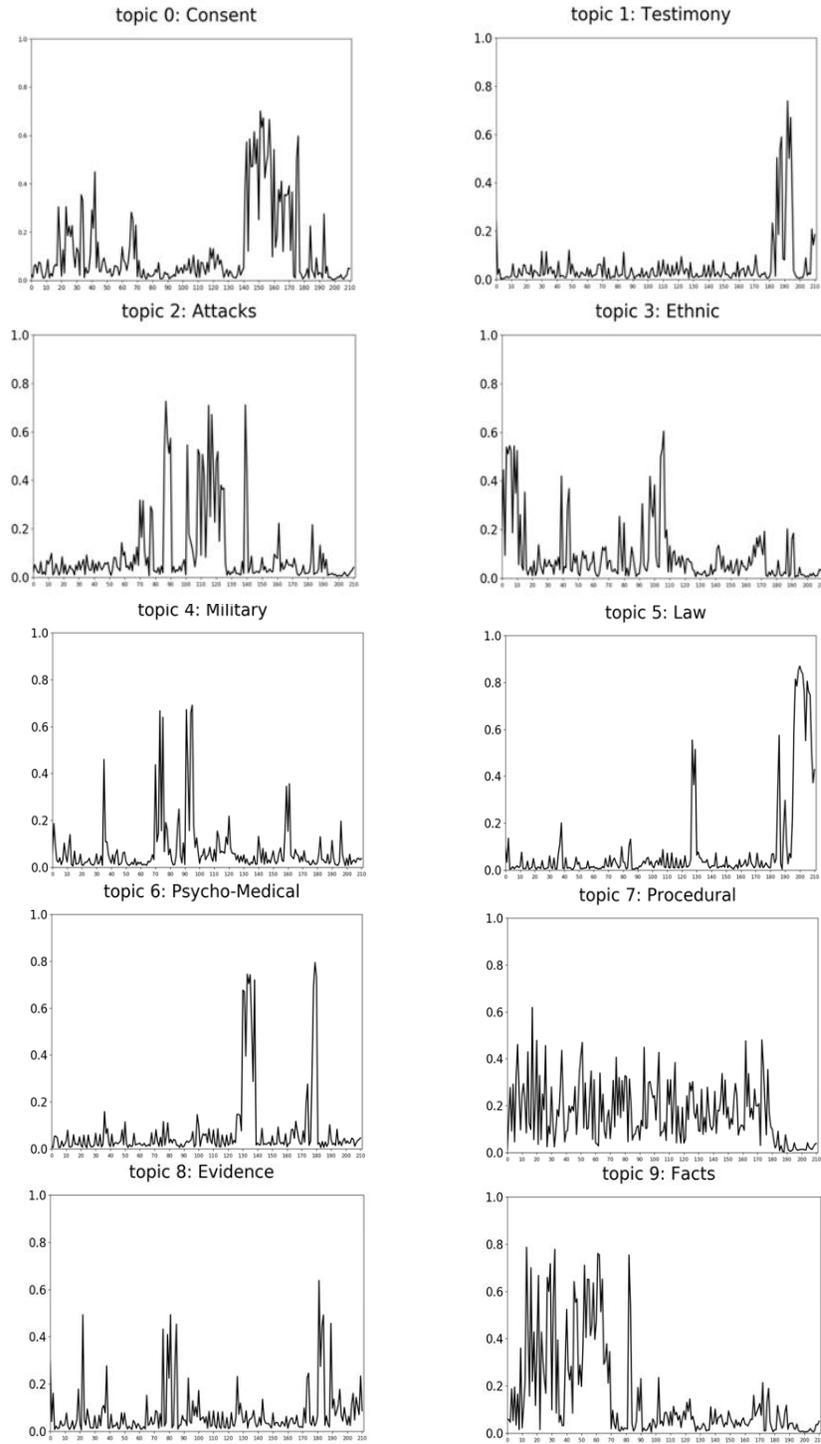
Up until now, we discussed the topics that emerged from the entire corpus. However, if we order the documents chronologically (with the judgment files grouped into a single timepoint), we can map the topics’ distribution over the course of the trial. Figure 5 shows a detailed analysis of each topic’s distribution chronologically. The graphs in figure 5 describe for each of the 10 topics, the topic’s frequency (Y axis) over the entire corpus arranged chronologically (X axis).¹⁰²

100. “FWS” is an abbreviation for Foča Witness Statement, which was used to identify some of the victims who gave testimony in the trial. Most witnesses, victims, and others, testified using pseudonyms and other protective measures.

101. U.N. Updated Statute of The International Criminal Tribunal For The Former Yugoslavia (Sept. 2009), https://www.icty.org/x/file/Legal%20Library/Statute/statute_sept09_en.pdf.

102. The transcripts were ordered chronologically from the prosecution’s opening statement on 20 March 2000 (=document no. 0) until the defense’s closing statement on 22 November 2000 (=document no. 184). The Judgment—rendered on 22 February 2001—was segmented according to the decision’s sections into 26 files (= document nos. 185–210; these files also included the transcript from the session of the oral reading of the judgment).

Figure 5: topic distribution over the course of the trial



Examining the topic distribution over the corpus helps convey when a certain topic was more prominent, or even introduced to the trial. Meta-trial topic 7 “procedural,” for example, which deals with trial procedure and administration appears quite consistently throughout the trial stages but is absent from the judgment itself. Topic 5 “law,” on the other hand, is mostly absent from the trial process itself and appears mainly in the judgment.

The chronological analysis of topic distribution is also pertinent to the case-specific topics: topic 9 “facts” is prominent in transcripts from the early stages of the trial, mostly the prosecution case. This makes sense, given that it is the prosecution’s burden to establish the factual framework.¹⁰³ Examining the two other topics relating to the factual framework, namely, topic 1 “testimony” and topic 8 “evidence” reveals, however, different patterns. The topic of evidence has several peaks throughout the trial, which reflects the way evidence is tendered, questioned, and examined all through the course of the proceedings, until its weight is determined in the judgment. The topic of testimony, on the other hand, is centered mainly in the documents of the judgment. One explanation may be that this topic does not describe the acts of testimonies themselves, namely the stories of the witnesses (that would be topic 9 “facts”), but rather the judgment’s interpretation of the testimonial accounts.

Against this background, it is interesting to see, for example, how topic 3 “ethnic tensions,” which belongs to the factual background of the case, is prominent both in the prosecution and the defense phases. However, it is downplayed in the judgment. This may signify the court’s reluctance to weigh in on the conflicting historical narratives about “who started the war,”¹⁰⁴ simply stating in the Findings section of the judgment that “an armed conflict between the Serb and Muslim forces broke out in Foča.”¹⁰⁵

D. Semantic Fields of Rape

It is not surprising that in a trial dedicated to adjudicating crimes of sexual violence, focusing on the “rape camps” in the Foča province in Bosnia, half of the topics will include the word “rape” (1, 5, 6, 8, 9), with the token “rape” appearing 2,330 times in the corpus.

Out of the five topics that explicitly include the term “rape,” three are tied to the fact-finding process: testimony, evidence and facts. The fourth one, psycho-medical, consists of terms relating to physical and psychological trauma of sexual violence. The fifth explicit invocation of the term “rape” is within the topic discussing the legal framework.

While a keyword search, for example, would have identified and returned all these results, what would have been lost is the disparate semantic contexts. Topic modeling is able to disambiguate the word-sense.¹⁰⁶ Word sense

103. Michail Vagias & Janos Ferencz, *Burden and Standard of Proof in Defence Challenges to the Jurisdiction of the International Criminal Court*, 28 LEIDEN J. OF INT’L L. 133, 155 (2015).

104. *Id.*

105. Marko Milanović, *Establishing the Facts about Mass Atrocities: Accounting for the Failure of the ICTY to Persuade Target Audiences*, 47 GEO. J. INT’L L. 1321, 1378 (2016).

106. Navigli, *supra* note 76.

disambiguation (WSD) is the task of determining the meaning of an ambiguous word in its context. It is an important problem in NLP.¹⁰⁷ Topic modeling captures the polysemic use of words and the different semantic fields and meanings.¹⁰⁸ In the case of the term “rape,” the model managed to capture at least three semantic fields dealing with rape: legal, psycho-medical, and factual.

Topic 0, which we named “consent,” requires some more illumination. Consent to the sexual act is the defining characteristic of rape. In *Kunarac*, the Trial Chamber expanded the existing *Furundžija* definition of consent (“relevant act of sexual penetration will constitute rape only if accompanied by coercion or force or threat of force against the victim or a third person”).¹⁰⁹ It concluded that for the purpose of the *actus reus* of rape in international law, consent “must be consent given voluntarily, as a result of the victim’s free will, assessed in the context of the surrounding circumstances.”¹¹⁰

Can the model reflect the conflicting narratives about consent and rape? The topic “consent” does not include the term “rape.” Its terms relate details regarding a relationship between the accused and other soldiers and certain girlfriends (Vukovic, Kovac, girl, girlfriend, soldier, café, coffee, apartment, flat). These words seem to reflect an alternative narrative for the sexual enslavement of Muslim girls by Serbian soldiers. This alternative version, which uses words like café, apartment, and girlfriend, is one describing a consensual relation rather than forced enslavement.¹¹¹ The term then, while not discussing rape, is in fact a mirror image, or a counter-narrative, of explicit “rape” related topics.

Judging from the model, the topic of consent is closely related to the defense narrative or the accused’s version of the events in question. This fits with the defense strategy in the trial which was either of complete denial of the acts or that the sexual relationship was consensual, as can be gleaned from the accused Kunarac’s testimony before the court:

Accused: Then she was sitting next to me. She fell on me. She put her head on my chest and begged me not to ask her anything. This gesture on her part did surprise me. . . . At that point, she started kissing me and begging me not to ask her a thing. At that moment, I was totally confused. I absolutely did not understand what she was doing. I tried to refuse this behaviour of her . . . After that, I accepted this behaviour of hers and we had full sexual vaginal intercourse . . . And I did not refuse her in any way; I accepted her behaviour.

JUDGE MUMBA: Is it your position, accused, that DB seduced you?

107. Jordan Boyd-Graber, David Blei & Xiaojin Zhu, “A Topic Model for Word Sense Disambiguation,” in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* (2007).

108. *Id.*

109. Prosecutor v. Dragoljub, Kunarac, Radomir, Kovic & Zoran Vukovic, Case No. IT-96-23-T, Judgement ¶ 436 (Feb. 22, 2001).

110. *Id.* at ¶ 460.

111. *Id.* at ¶ 75.

Accused: I didn't say I wanted it. At that moment, I had sexual intercourse with her against my will. I mean, without having a desire for sex. ...

JUDGE MUMBA: That's enough. Questions, please.¹¹²

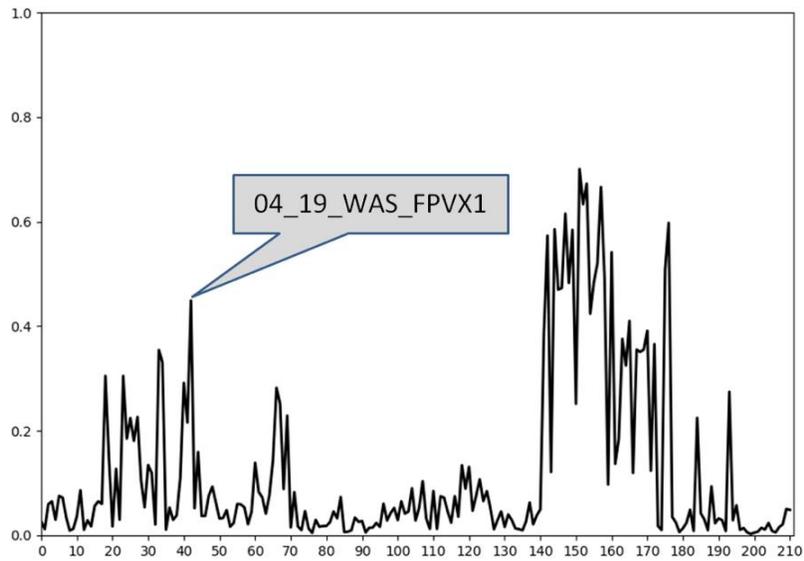
Examining the topic distribution over the course of the trial, with emphasis on the semantic fields of rape, demonstrates the potential of topic models not only to identify the themes of the corpus, but also to distinguish between the competing narratives of the adversaries in the trial—the prosecution and the defense. The model's strength is in exposing these competing narratives as two different topics. The relative absence of the topic of consent from the judgment, reflects the Trial Chamber's categoric rejection of the defense's version regarding the consensual nature of the sexual acts.

Examining the chronological distribution of topic 0 “consent” provides an additional support for the hypothesis that the topic captures the Accused version. Figure 6 shows that the topic of consent is most prominent in the latter part of the trial, namely the defense phase. Yet, the first part of the trial, generally the prosecution phase, also has several areas in which the topic is frequent. This raises a question as to the presence of the “counter-narrative” of the accused within the prosecution narrative. Since each point on the X axis stands for a document, we can trace those specific documents that have a high percentage of the topic “consent” in them within the prosecution phase. The meta-data coded in the caption “04_19_WAS_FP VX1”¹¹³ indicates that this instance in the trial, which is during the prosecution case but is unexpectedly characterized by words relating to the topic which we tagged as “consent” (topic 0), stems from a cross-examination by the defense of a prosecution witness who testified about crimes of sexual violence that were perpetrated against her. This serves to explain the co-presence of two narratives—that of the prosecution and that of defense. This movement from the bird's-eye overview to the zooming into a specific moment in the trial is one of the key achievements of employing topic modeling in the legal domain. Fig 6 demonstrates the potential of topic modeling analysis to identify and reveal the battle of narratives between the prosecution and the defense over the key legal question in the trial: the question of rape.

112. Prosecutor v. Dragoljub, Kunarac, Radomir, Kovic & Zoran Vukovic, Case No. IT-96-23-T, Transcript, p. 4541–42 (July 6, 2000).

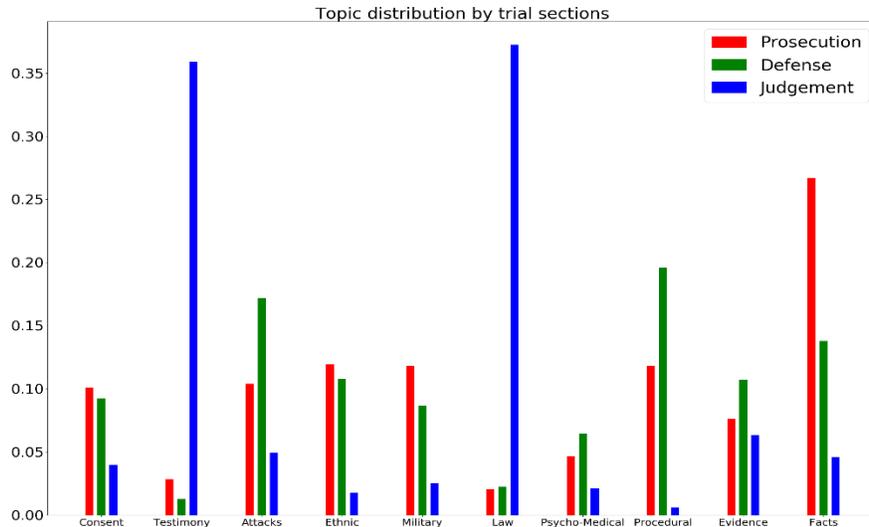
113. Fields indicating date of testimony (April 19), witness ID (WAS), witness gender (Female), witness called by (Prosecution), witness type (Victim) and witness questioning (X = Cross examination, part 1).

Figure 6: Narrative of Rape and Consent



As we mentioned earlier, each transcript was tagged for its metadata, including the witness function (expert, fact, victim, accused), if it is a prosecution or defense witness and if it is direct or cross examination. Based on this indexing, we grouped the documents into two categories: prosecution narrative and defense narrative. Differently from the chronological distribution of topic in Figure 4, in the current bar chart, cross examination of a defense witness by the prosecutor, for example, is included in the prosecution section, despite the fact that chronologically it was conducted in the latter half of the trial, during defense stage. Figure 7 shows the bar chart of the topic distribution according to trial sections.

Figure 7: topic distribution by trial sections



The most prominent topic in the prosecution case is topic 9 “facts” in 25% of the prosecution corpus. For the defense, it is the procedural topic, topic 7 that is most frequent, in 20% of the defense corpus. Topic 2 “attacks on civilian population” appears in 17% of the documents.

Against the more dispersed topical distribution of the parties’ narratives, two main topics comprise the judicial opinion: legal framework and testimony, each more than 35%. The judgment then exhibits coherent topical composition, with some of the other topics, such as ethnic tension or military structure hardly reflected in the judicial corpus. More than whether or not there is a fully formed prosecution or defense narrative, what the topic model shows is that the judgment exhibits a strong narrative structure.

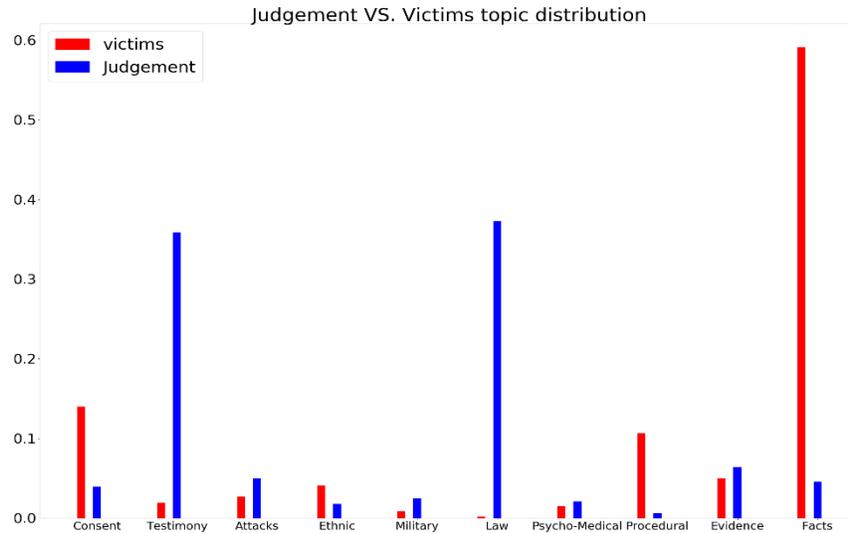
E. Assessing the Gaps between Victims’ Testimony and Judicial Opinion

The discussion thus far has shown topic modeling to be a reliable empirical tool to parse through trial transcripts and extract the hidden themes. This makes it a useful exploratory tool to be utilized both during the legal process, by judges and counsels, as well as in the critical analysis of the trial proceedings by legal scholars.

In this subsection, our aim is to indicate, albeit suggestively, the potential of the topic modeling technique to serve as an ex-post evaluation of the court’s attentiveness to victims’ narratives by identifying the gaps between the topics emerging from the victims’ testimonies and those that comprise the judgment. For this purpose, we isolated the files containing only testimonies of victims, and only direct examination. Figure 8 presents the comparison between the topic distribution of the judgment and the topic distribution of the victims’ transcripts.

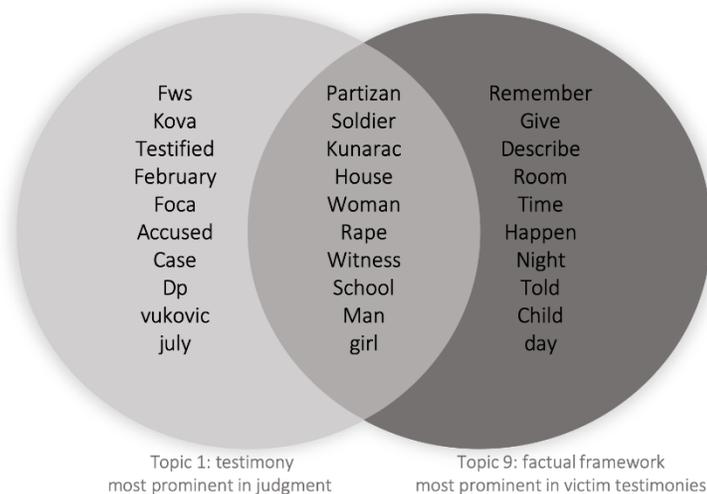
As we mentioned before the judgment consists of two prominent topics, the “testimony” topic and the “legal framework” (topic 1 and 5 respectively). Compared to that, the victims’ narrative strongest topic is topic 9, the “facts” (60%).

Figure 8: topic distribution of victim testimony vs. judgment



What can be learned from the disparities between the victims’ stories and the judgment if we compare the two fact-finding related topics: “facts” (victims) and “testimony” (judgment) that dominate the respective corpora? Figure 9 compares the 20 top ranked terms in each topic. It reveals that in fact there are significant overlaps and similarities between the two topics. While the Venn diagram does not take into account the different weight of the term in each topic, it does give us a visual aid to see more concretely the thematic content of each topic. The nouns “house,” “school,” “soldier,” “man,” and the name of the accused Kunarac, for example, appear in both topics.

Figure 9: topic 1 "testimony" and topic 9 "facts" compared



Where the topics differ however is in the verbs: topic 1 “testimony” has a single verb, “testified,” while a quarter of the top 20 terms in topic 9 “facts” are verbs: “remember;” “happen;” “told;” “give;” “describe.” These verbs express modes of agency and action, that characterize the narrational act of giving testimony. They describe both the events which the testimony relates (happen, told) and also the act of giving testimony (remember, describe).

The topic composition comparison indicates that both topics discuss similar factual events, expressed in their overlapping noun composition, but reflect different temporal perspective and narrative points of view on the events related.

The Kunarac case is a landmark decision in combating conflict related sexual violence. It is a noteworthy judgment, which respects the victims and delivers justice. Nevertheless, in the Kunarac trial, as in many other international trials, the Prosecution faced grave evidentiary hurdles in proving the crimes.¹¹⁴ The tension ultimately rests on bridging the gap between the widespread, systematic campaign of “ethnic cleansing” and mass rape whose horrific details were related in the court and the evidentiary challenges in proving concrete incidents of rape based on the multitude of personal experience related by the victims:

By their very nature, the experiences which the witnesses underwent were traumatic for them at the time, and they cannot reasonably be expected to recall the minutiae of the particular incidents charged, such as the precise sequence, or the exact dates and times, of the events they have described. The fact that these witnesses were

114. Gideon Boas, *Creating Laws of Evidence for International Criminal Law: The ICTY and the Principle of Flexibility*, 12 in CRIM. L. FORUM 41, 90 (2001).

detained over weeks and months without knowledge of dates or access to clocks, and without the opportunity to record their experiences, only exacerbated their difficulties in recalling the detail of those incidents later. In general, the Trial Chamber has not treated minor discrepancies . . . as discrediting their evidence where that witness has nevertheless recounted the essence of the incident charged in acceptable detail. . . .

The Trial Chamber has also taken into account the fact that these events took place some eight years before the witnesses gave evidence. . . .

Many of these witnesses were minors at the time of the events which they described, some of them as young as fifteen years. The level of detail which such witnesses could be expected to recall is different to that expected of witnesses who were more mature at the relevant time.¹¹⁵

As one trial commentator wrote, the decision in *Kunarac* has “an air of unreality” as it relates horrible experiences of abuse and suffering, at the same time as those experiences are measured against convoluted and sometimes archaic rules of evidence and relevance.¹¹⁶ In *Kunarac*, for example, the decision runs to over 300 pages and contains masses of evidence and testimony on the experience of women raped and abused while in detention.¹¹⁷ In evaluating this evidence, the Tribunal detailed individual acts of abuse, noting times, places, supporting evidence.¹¹⁸ On a number of indictments, the Tribunal ruled that there is insufficient evidence, or the indictment has not been properly drawn, to sustain a conviction.¹¹⁹ This creates a gap between the quantity of testimony and the quality of evidence. As Buss summarizes, “[a]fter reading pages and pages of witness evidence about the existence of what are, in effect, ‘rape camps,’ the Tribunal then dismisses individual charges of rape for lack of evidence. Arguably, within a strict reading of criminal evidence and international law, the Tribunal is “right” in its conclusion, but the very process of analysis feels particularly ill-suited to the task of accounting for, and reconciling the experiences of these women.”¹²⁰ Buss asks with genuine puzzlement, “[w]hat lesson do we take away from a Tribunal that can, on the one hand, find evidence of rape camps, and yet find that individual women are not to be believed in their accounts of rape and violence at those same camps?”¹²¹

Buss’ grim criticism revisits the frustrating gap with which this article opened between the quantity of voices and the quality of evidence that plagues international criminal trials in general and trials dealing with systematic sexual

115. Prosecutor v. Kunarac, Case No. IT-96-23-T & IT-96-23/1-T, Judgment, ¶¶ 564–65 (Int’l Crim. Trib. for the Former Yugoslavia Feb. 22, 2001).

116. Buss, *supra* note 14, at 98.

117. Kunarac, Case No. IT-96-23-T & IT-96-23/1-T at ¶ 574.

118. *Id.*

119. Buss, *supra* note 14, at 98.

120. *Id.*

121. *Id.* at 99.

violence in particular.¹²² Nevertheless, examining the results of the topic model may mean that the semantic gap between the victim's stories and the court's judgment might not be that big after all. The empirical data shows that there is a high degree of overlap between top ranked terms in the two most frequent topics in the victims' narratives and in the judgment. This may indicate the court's attentiveness to the stories of the victims.

This encouraging result does not mean that we reached utopia. Far from it. It may, however, help dispel unnecessary institutional anxieties over quantity, allowing for a more careful consideration of the functions and potential contribution of mass testimony to the understanding of mass atrocity.¹²³ What the model does is to provide preliminary empirical data to direct our practical and scholarly efforts in the right direction. If the court is capable of listening to large number of witnesses and victims, then perhaps the gap is not one of "miscommunication," but rather of "admissibility," namely what the courts are "allowed" to hear in the first place, which is a function of tribunals' evidentiary regimes.

V. CONCLUSION

In this article, we presented the results of a novel method for engaging with large testimonials corpora, applying a generative statistical model known as *unsupervised topic modeling* to a corpus containing testimonies of witnesses, victims and perpetrators of conflict-related sexual violence before international criminal tribunals. Employing an LDA-based topic modeling, we explored latent themes and semantic fields in the trial corpus. In particular, we highlighted the potential of topic modeling methods, rooted in Digital Humanities, to overcome critical impediments in empirical legal studies.

We showed how the model classifies large amounts of textual data into semantically similar clusters without human direction and, crucially, without reducing the documents into binary or numerical form, making it especially attractive to the analysis of complex narratives and free-form texts, such as court transcripts, and testimonies in particular.

Focusing on trial transcripts from the ICTY's Kunarac case, dealing with systematic mass rape of Muslim girls and women by Bosnian-Serb forces, the article demonstrated 1) LDA algorithm's capacity to provide a reliable view of the trial through a topic modeling of its transcripts, and 2) the potential of topic modeling to illuminate gaps in the themes emerging from the legal process, reflected in the court transcripts—in particular victim testimony—and the judicial opinion expressed in the judgment. Taken together, the article demonstrated the potential contribution of topic modeling to the field of empirical legal studies, both as a practical heuristic mechanism that can be

122. *See id.* ("What lesson do we take away from a Tribunal that can, on the one hand, find evidence of rape camps, and yet find that individual women are not to be believed in their accounts of rape and violence at those same camps? In this respect, the Yugoslav Tribunal reproduces many of the same problematic aspects of 'the rape trial' found in Western domestic legal systems.").

123. Keydar, *supra* note 6, at 518.

employed during the legal proceeding, and in its ex-post analysis in legal scholarship.

The article's contribution is in bridging the gap between the field of Digital Humanities and its increasing use of topic models for analyzing literary texts, and other text-based disciplines, in particular the law. Topic modeling method can assist scholars and practitioners in making sense of complex legal cases, involving large amounts of testimonies, documents and data, while preserving the vocabulary of the individual witness.

The introduced method might also be a means for judges to organize the vast quantities of information that is said, but perhaps not always heard, over the course of the trial. It may be thought of as a form of quality control mechanism for judicial decision-making, during or after the trial. This is not to say that an adoption of a topic modeling approach will make judgments more accurate than they are at present, or that judges who will employ this method will enjoy a full grasp of the case at hand. Moreover, topic model-based "distant listening" does not replace careful, attentive close reading of testimonies. Rather, it is best thought of as amplifying and augmenting careful reading and thoughtful analysis. Instead of thinking of the computer as replacing the human, the ideal is to combine machine-enabled distant reading with more traditional interpretative techniques of close reading.

The strength of the method demonstrated in the article is its applicability to the broader legal field. Future directions emerging from this research include: 1) enlarging the analyzed database: topic modeling has the potential to enable cross-tribunal comparisons of the themes appearing in testimonies in relation to similar offences, in different locales, cultures and institutional settings—in particular different evidentiary and procedural regimes; 2) broadening the scope of the analysis beyond international tribunals to other legal arenas that face complex litigation cases; 3) topic modeling could help illuminate a range of behavioral questions relating to judicial attentiveness in complex legal litigation and institutional design, as a function of the courts' diverse evidentiary and procedural regimes.