

Processing legal text: Background and challenges

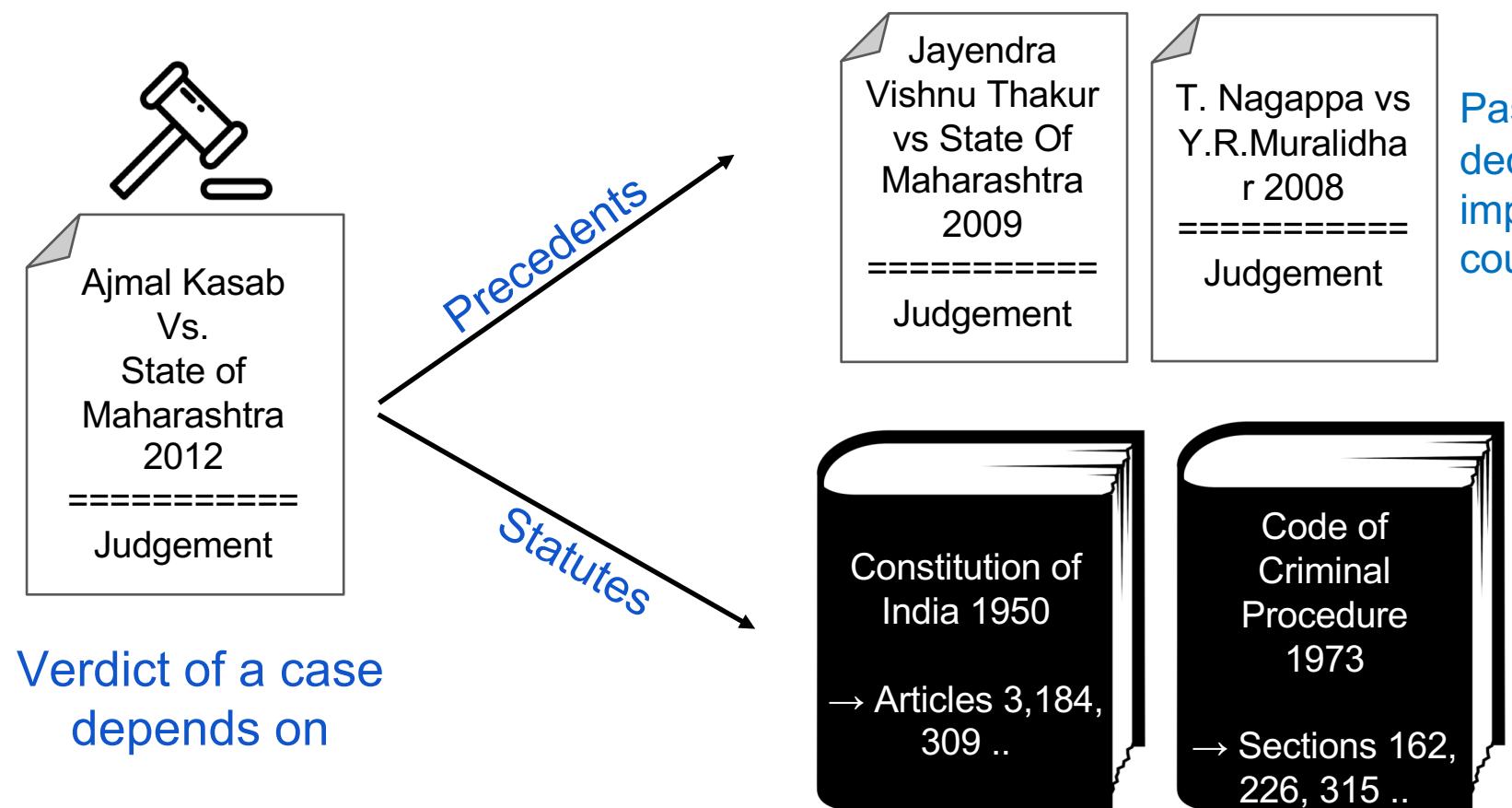
Saptarshi Ghosh

Department of CSE
Indian Institute of Technology Kharagpur
<http://cse.iitkgp.ac.in/~saptarshi>

Types of Law systems

- Broad types of Law systems – Civil Law and Common Law
- **Civil Law**
 - Law is defined by **codified statutes**, ordinances, etc.
 - Followed in most of Europe and South America
- **Common Law**
 - Law is defined by prior judgements pronounced by (important) courts
 - **Stare decisis**: similar facts or similar situations should yield similar results
- Many countries follow a combination of Criminal Law with Civil Law
 - Followed in USA, UK, India, Canada, Australia, ...

Two sources of law in a Common + Civil Law system



Types of legal documents

Many types of legal documents

- Statutes, e.g., Acts, Sections
- First Information Reports (FIR)
- Docs generated during a case, intermediate orders, ...
- Case judgements (generated at the end of a case)
- Contracts, MoUs
- Patents
- Law reports
- ... any many more

Many types of legal documents

- **Statutes, e.g., Acts, Sections**
- First Information Reports (FIR)
- Docs generated during a case, intermediate orders, ...
- **Case judgements (generated at the end of a case)**

- Contracts, MoUs
- Patents
- Law reports

- ... any many more

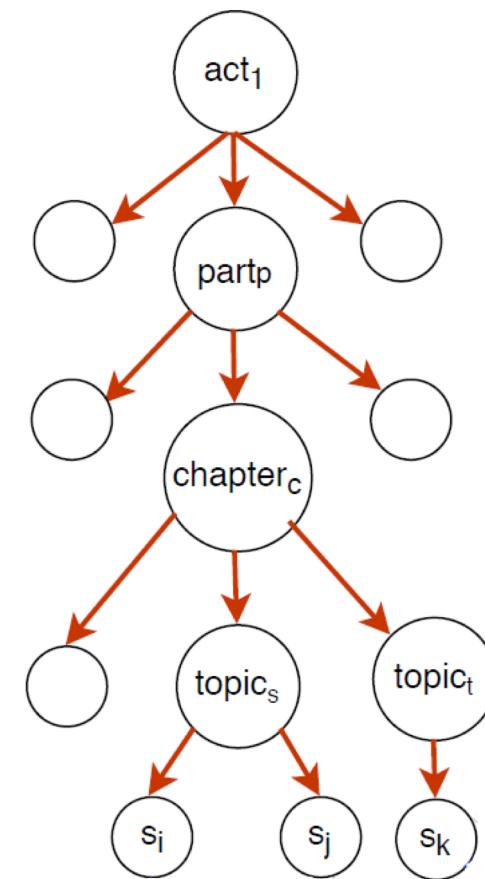
Focus of this tutorial
Usually publicly available
Lot of prior research

Statutes

- Laws are usually encoded in “Acts”
 - Road Traffic Act, 1961 – Ireland
 - Immigration and Asylum Act, 1999 – UK
 - European Convention on Human Rights (ECHR), 1950 – EU
 - Indian Penal Code (IPC), 1860 - India
 - Homeland Security Act, 2002 – USA
- New laws are formed regularly
- Older laws are extended / amended

Statutes

- Acts usually divided into sub-parts
 - Parts, Chapters, Topics, Sections
 - A tree-like hierarchy
 - Lowest level – Sections / Articles
 - Similar sections grouped under a topic
 - Similar topics grouped under a chapter
 - ...



Example: Immigration and Asylum Act 1999 (UK)

- “An Act to make provision about immigration and asylum; to make provision about procedures in connection with marriage on superintendent registrar’s certificate; and for connected purposes.”
- Divided into 10 Parts
- Each Part contains multiple related Sections

Part I Immigration: General

Part II Carriers' Liability

Part III Bail

Routine bail hearings

44. Bail hearings for detained persons.
45. Location of bail hearings.
46. General right to be released on bail.
47. Powers exercisable on granting bail.
48. Forfeiture.
49. Forfeiture of securities.
50. Power of arrest.

Procedure

51. Procedure.
52. Use of live television links at bail hearings.

Bail hearings under other enactments

53. Applications for bail in immigration cases.
54. Extension of right to apply for bail in deportation cases.

Grants

55. Grants to voluntary organisations.

Part IV Appeals

Part V Immigration Advisers and Immigration Service Providers

Part VI Support for Asylum-Seekers

Part VII Power To Arrest, Search and Fingerprint

Part VIII Detention Centres and Detained Persons

Part IX Registrar's Certificates: Procedure

Part X Miscellaneous and Supplemental

Example: European Convention on Human Rights (EU)

- Protects the human rights of people in EU countries and prohibits unfair and harmful practices.
- Divided into several Articles (which actually state the laws) which are grouped into Sections
- Extended through multiple Protocols
 - the right to life (Article 2)
 - freedom from torture (Article 3)
 - freedom from slavery (Article 4)
 - the right to liberty (Article 5)
 - the right to a fair trial (Article 6)
 - the right not to be punished for something that wasn't against the law at the time (Article 7)
 - the right to respect for family and private life (Article 8)
 - freedom of thought, conscience and religion (Article 9)
 - freedom of expression (Article 10)
 - freedom of assembly (Article 11)
 - the right to marry and start a family (Article 12)
 - the right not to be discriminated against in respect of these rights (Article 14)
 - the right to protection of property (Protocol 1, Article 1)
 - the right to education (Protocol 1, Article 2)
 - the right to participate in free elections (Protocol 1, Article 3)
 - the abolition of the death penalty (Protocol 13)

Example: Indian Penal Code (IPC)

- Collection of statutes that deal with major types of crimes in India
 - Divided into Chapters, Topics, Sections
 - Similar sections grouped under a topic
 - Similar topics grouped under a chapter

Chapter	Topic	Section
Offences affecting Human Body	Offences affecting Life	299: Culpable homicide
		307: Attempt to murder
	Hurt	321: Voluntarily causing hurt
		334: Voluntarily causing hurt on provocation
Offences against Property	Robbery and Dacoity	390: Robbery
		396: Dacoity with murder
	Criminal Trespass	441: Criminal trespass 446: House breaking by night

Statutes

- Laws are usually coded in “Acts”
 - Road Traffic Act, 1961 – Ireland
 - Immigration and Asylum Act, 1999 – UK
 - **European Convention on Human Rights (ECHR), 1950 – EU**
 - **Indian Penal Code (IPC), 1860 - India**
 - Homeland Security Act, 2002 – USA

The ECHR Articles and IPC Sections are used in two popular Legal NLP datasets for **Legal Statute Identification (LSI)**

LSI - given a situation, which statutes may have been violated
– multi-class, multi-label classification

Court case judgements

- The final judgement pronounced by a Court of Law at the end of a case
- Contains metadata (preamble) and the main judgement text
- Judgements by important courts usually publicly available, particularly in Common Law jurisdictions (caselaw)
 - <https://www.echr.coe.int/Pages/home.aspx?p=caselaw&c> - ECHR
 - <https://main.sci.gov.in/judgments> - India
 - <https://www.supremecourt.uk/decided-cases/> - UK

The screenshot shows the homepage of the European Court of Human Rights (ECHR). At the top, there is a banner featuring the court's name in English and French, along with the Council of Europe logo and a search bar. Below the banner, a navigation menu includes links for Home, The Court, Case-law (which is highlighted in blue), Press, Hearings, Statistics, Applicants, Official texts, Events, Library, and Français. A large blue banner with yellow stars from the European Union flag spans across the page. Below this, a section titled "HUDOC database" is visible, featuring the HUDOC logo and a brief description of the database's content.

European Court of Human Rights
Cour européenne des droits de l'homme

Council of Europe | ECHR-KS | Search HUDOC

Search ECHR website

Home The Court Case-law Press Hearings Statistics Applicants Official texts Events Library Français

Judgments and decisions

HUDOC database

All the Court's judgments and a large selection of decisions, information on communicated cases, advisory opinions, press releases, legal summaries and Commission decisions and reports are published in the HUDOC database.

More information

A case judgement from INSC

- Decided by the Supreme Court of India on 3 May 2013
- Metadata / preamble
 - Decided by which court
 - Date on which judgement was pronounced
 - The contending parties (Appellant vs. Respondent)
 - Name of the judge / bench
 - Sometimes, names of lawyers representing the two parties

Ankush Shivaji Gaikwad vs State Of Maharashtra on 3 May, 2013

Author:J.

Bench: T.S. Thakur, Gyan Sudha Misra

REPORTABLE

IN THE SUPREME COURT OF INDIA
CRIMINAL APPELLATE JURISDICTION
CRIMINAL APPEAL NO. OF 2013
(Arising out of S.L.P. (Crl.) No.6287 of 2011)

Ankush Shivaji Gaikwad ...Appellant

Versus

State of Maharashtra ...Respondent

JUDGMENT

T.S. THAKUR, J.

1. Leave granted.
2. This appeal arises out of a judgement and order dated 24th August, 2010 passed by the High Court of Judicature at Bombay, Aurangabad Bench, whereby Criminal Appeal No.359 of 2008 filed by the appellant and two others has been dismissed in so far as the appellant is concerned and allowed qua the remaining two, thereby upholding the appellant's conviction for the offence of murder punishable under [Section 302](#) of the I.P.C and the sentence of imprisonment for life with a fine of Rs.2,000/- awarded to him. In default of payment of fine the appellant has been sentenced to undergo a further imprisonment for a period of three months.
3. The factual matrix in which the appellant came to be prosecuted and convicted has been set out in detail by the trial Court as also the High Court in the orders passed by them. We need not, therefore,

The main text of a case judgement contains:

- **Facts**: chronology of events that led to filing the case
- **Ruling by Lower Court**: judgements given by lower Courts
- **Arguments**: arguments of the contending parties
- **Statute citations**: citations to relevant laws / statutes
- **Precedent citations**: citations to prior cases
- **Ratio of the decision**: rationale of the Court for the final ruling
- **Final Judgement**: conclusion of the Court

Rhetorical
segments

Example case: Case of Z.A. v. Ireland, European Court of Human Rights

(Application no. [19632/20](#))

- **Facts**
- **Ruling by Lower Court**
- **Arguments**
- **Statute citations**
- **Precedent citations**
- **Ratio of the decision**
- **Final Judgement**

THE FACTS

Criminal proceedings
Deportation proceedings

High Court judgment of 30 May 2019
Decision of 13 Nov 2019 of the Supreme Court

RELEVANT LEGAL FRAMEWORK AND PRACTICE

RELEVANT STATUTORY PROVISIONS
RELEVANT CASE-LAW

THE LAW

ALLEGED VIOLATION OF ARTICLE 8 OF THE CONVENTION

The parties' arguments

The Court's assessment

ALLEGED VIOLATION OF ARTICLE 13 OF THE CONVENTION

The parties' arguments

The Court's assessment

FOR THESE REASONS, THE COURT, UNANIMOUSLY ...

A challenge ... Case judgements across countries not consistent

- Case judgements from all countries usually have these rhetorical segments
 - Some segments may be missing in very short judgements
- But writing style can vary hugely across countries / courts
 - Segments can be interleaved, e.g., facts often get interleaved with ruling by lower court, arguments
 - Headings can be worded very differently
 - There can be no headings at all !!

Example case 2

United Kingdom Supreme Court

The Manchester Ship Canal Company Ltd and another (Respondents) v United Utilities Water Plc (Appellant)

2014

LORD SUMPTION (with whom Lord Clarke and Lord Hughes agree)

Introduction

1. The question at issue on this appeal is whether a sewerage undertaker under the Water Industry Act 1991 has a statutory right to discharge surface water and treated effluent into private watercourses such as the Respondents' canals without the consent of their owners.
2. Discharge into a private watercourse is an entry on the owner's land, and as such is an unlawful trespass unless it is authorised by statute. It is common ground that no express statutory right is conferred by the Water Industry Act. The question is therefore whether it should be implied. A statutory right to commit what would otherwise be a tort may of course be implied. But since this necessarily involves an interference with the rights of others, the test has always been restrictive. The implication must be more than convenient or reasonable. It must be necessary. As a general rule, this will involve showing either that the existence of the power is necessarily implicit in the express terms of the statute, or else that the statutory purpose cannot be effectually achieved without the implication. In particular a right to commit what would otherwise be a tort may be implied if a statutory power is incapable of being exercised or a statutory duty is incapable of being performed without doing the act in question: *Manchester Corporation v Farnworth* [1930] AC 171, 183 (Viscount Dunedin), *Allen v Gulf Oil* [1981] AC 1001, 1013 (Lord Wilberforce).

The law before 1991

3. It has been said that a court "should not routinely investigate the statutory predecessors of provisions in a consolidation statute": *R v Secretary of State for the Environment, Transport and the Regions, Ex p Spath Holme Ltd* [2001] 2 AC 349, 388 (Lord Bingham of Cornhill). This is not so much a rule of construction as a valuable warning against the over-ready assumption that a consolidating Act means exactly the same as the enactments which it replaces. There are, however, cases where a consolidating Act cannot be understood without reference to the state of the law as it was when it was enacted. This is one of them.

At para 201, Lord Rodger suggested that the test could well be expressed thus:

"would the consequences of applying the statutory provision retroactively, or so as to affect vested rights or pending proceedings, be 'so unfair' that Parliament could not have intended it to be applied in these ways? In answering that question, a court would rightly have regard to the way the courts have applied the criterion of fairness when embodied in the various presumptions."

The first question

57. So far as the first question is concerned, Mr Karas QC, on behalf of United Utilities, a sewerage undertaking, relied primarily to support his case for a positive answer, on the provisions of section 159 of the 1991 Act. I would reject that case and there is nothing I can usefully add to what Lord Sumption and Lord Toulson say in paras 13-15 and 26-28 of their respective judgments. At least in relation to sewers laid after the 1991 Act came into force, United Utilities' argument is not supported by the language of section 159 or any other provision of the 1991 Act. It is inconsistent with some other provisions of the 1991 Act, and it is not supported by any practical considerations (although it is fair to add that it is not undermined by any practical considerations either). The reasoning of all three members of the Court of Appeal in *British Waterways Board v Severn Trent Water Ltd* [2002] Ch 25, summarised in para 14 above, appears to me to be unanswerable.

The second question

58. The question whether sewerage undertakers can claim any rights in respect of any outfalls must ultimately turn on the 1991 Act, but in my view, the issue should be addressed by first identifying the water authorities' rights in respect of outfalls from public sewers immediately before the 1989 Act came into force. Mr Karas's argument is that it is a necessary inference

Example Case 3

Supreme Court of India

Subimal Sarkar v Sachindra Nath Mondal and others, 2003

1. Original complainant in Sessions Case No.127 of 1989 before the Sessions Judge, West Dinajpur, at Balurghat is in appeal before us in the above Criminal Appeal, against the judgment of acquittal made by the High Court Calcutta in Criminal Death Reference No.4/1990 and Criminal Appeal No.327 of 1990. The State has not preferred any appeal against the judgment of the High Court, but is a party respondent before us.

2. Prosecution case briefly stated is that one Suchitra, the daughter of the appellant herein was married to Nakul Chandra son of Sarat Chandra Mandal, accused No.1 before the Sessions Court. At the time of the marriage, there was an agreement to pay dowry in cash of Rs. 500/- out of which, Rs. 300 was paid to the accused A-1 by the appellant but he could not pay the balance amount of dowry due to poverty. A-1 was living with his wife A-4 and five of his children of which A-3 Sachindra Nath Mandal is one of them who was living with his wife Gauri Mandal, who was A-4 before the Sessions Court. The husband of deceased Suchitra was Nakul Mandal, who was also staying with his father. It is the prosecution case that because of the non-payment of balance of dowry, there was constant torture and ill-treatment of the deceased by the accused persons. This was made known to the appellant (PW-11) as also his wife PW-1 Arati Sarkar.

3. On 16.8.1986, PW-1 had come to know that her daughter had consumed poison in the house of accused No.1 (father in-law) and in view of the fact PW-11, husband of PW-1 was along with PW-2 away in the field, had proceeded towards the village of her sambandhi and on the way they met one Nakul Bhunia, who took them on his cycle towards Durlavpur. Further on the way, they found the body of deceased Suchitra being carried on Plank tied to the same by A-3 Sachinder Mandal with the help of some people and on seeing them A-3 and others left the body there and went to some distance. When PW-1 and PW-2 went near the body, they found bleeding from the mouth and nostril of the deceased, therefore they took the deceased across the river Pagliganj to Balurghat Hospital, where PW-15 Dr. Nath examined the deceased and declared her as brought dead. He also opined that, the death was caused by throttling. On the basis of the information, received from the Dr. PW-15, the police of the Balurghat Police Station registered a case under Section 302 [IPC](#) and started the investigation. The body was sent for post-mortem examination, which was conducted by Dr. D. Shah PW-14 on 17.8.1986. He opined that cause of death of the deceased was due to manual strangulation and was homicidal in nature. During the post-mortem, he noticed the bruise mark on the neck and both on right and left side of the wind pipe. On dissection, he found the hyoid bone fractured. After completion of the investigation, the police filed charge-sheet against the above-mentioned four accused persons.

4. The prosecution has primarily relied upon the evidence of PW-1, the mother, PW-2 aunt, PW-4 brother, PW-11, the father and another relative PW-2, Govind Sarkar to establish the fact that there was a dowry demand, pursuant to which the deceased was being tortured and on the date of incident the body of the deceased was being carried on a plank tied to the same. The prosecution also relied upon the medical evidence to establish the fact that there was a homicidal death by throttling.

5. A very peculiar defence was taken by the accused persons who contended that on the fateful day, the deceased had consumed poison and when it was noticed by her husband, Nukul, he tried to prevent the deceased from swallowing the poison by pressing the neck at which time A-3, brother in law of the deceased allegedly came and put his finger into the throat of the deceased to make her vomit the poison. The further contention was that at that point of time, deceased was alive, therefore, they decided to take her to the hospital and it was during that time that the mother of the deceased met on the way when the body was being taken to the hospital. The Trial Court on consideration of the evidence on record came to the conclusion that the prosecution by circumstantial evidence has established the case against the accused, hence held the accused persons guilty and awarded capital punishment to A-3, the brother in law of the deceased, Suchindra Nath Mandal, while the other three accused persons were convicted under Section 302 read with 34 and were awarded life imprisonment.

1. Original complainant in Sessions Case No.127 of 1989 before the Sessions Judge, West Dinajpur, at Balurghat is in appeal before us in the above Criminal Appeal, against the judgment of acquittal made by the High Court Calcutta in Criminal Death Reference No.4/1990 and Criminal Appeal No.327 of 1990. The State has not preferred any appeal against the judgment of the High Court, but is a party respondent before us.

No headings or divisions

2. Prosecution case briefly stated is that one Suchitra, the daughter of the appellant herein was married to Nakul Chandra son of Sarat Chandra Mandal, accused No.1 before the Sessions Court. At the time of the marriage, there was an agreement to pay dowry in cash of Rs. 500/- out of which, Rs. 300/- was paid to the accused A-1 by the appellant but he could not pay the balance amount of dowry due to poverty. A-1 was living with his wife A-4 and five of his children of which A-3 Sachindra Nath Mandal is one of them who was living with his wife Gauri Mandal, who was A-4 before the Sessions Court. This marriage 12 December 1981 was Nakul 11/11/81, who was 18 years old at that time. In the execution of the marriage because of the non-payment of balance of dowry, there was constant torture and ill-treatment of the deceased by the accused persons. This was made known to the appellant (PW-1) as also his wife PW-1 Arati Sarkar.

3. On 16.8.1986, PW-1 had come to know that her daughter had consumed poison in the house of accused No.1 (father in-law) and in view of the fact PW-11, husband of PW-1 was along with PW-2 away in the field, had proceeded towards the village of her sambandhi and on the way they met one Nakul Bhunia, who took them on his cycle towards Durlavpur. Further on the way, they found the body of deceased Suchitra being carried on Plank tied to the same by A-3 Sachinder Mandal with the help of some people and on seeing them A-3 and others left the body there and went to some distance. When PW-1 and PW-2 went near the body, they found bleeding from the mouth and nostril of the deceased, therefore they took the deceased across the river Pagliganj to Balurghat Hospital, where PW-15 Dr. Nath examined the deceased and declared her as brought dead. He also opined that the deceased was strangled. On the basis of the information, received from the Dr. PW-15, the police of the Balurghat Police Station registered a case under Section 302 IPC and started the investigation. The body was sent for post-mortem examination, which was conducted by Dr. D. Shah PW-14 on 17.8.1986. He opined that cause of death of the deceased was due to manual strangulation and was homicidal in nature. During the post-mortem, he noticed the bridle marks on the neck and both the right and left side of the wind pipe. On dissection, he found the hyoid bone fractured. After completion of the investigation, the police filed charge-sheet against the above-mentioned four accused persons.

4. The prosecution has primarily relied upon the evidence of PW-1, the mother, PW-2 aunt, PW-4 brother, PW-11, the father and another relative PW-2, Govind Sarkar to establish the fact that there was a dowry demand, pursuant to which the deceased was being tortured and on the date of incident the body of the deceased was being carried on a plank tied to the same. The prosecution also relied upon the medical evidence to establish the fact that there was a homicidal death by throttling.

5. A very peculiar defence was taken by the accused persons who contended that on the fateful day, the deceased had consumed poison and when it was noticed by her husband, Nukul, he tried to prevent the deceased from swallowing the poison by pressing the neck at which time A-3 (brother in law) of deceased George D. Shah and others forced him to the foot of the deceased to make her vomit the poison. The further contention was that at that point of time, deceased was alive, therefore, they decided to take her to the hospital and it was during that time that the mother of the deceased met on the way when the body was being taken to the hospital. The Trial Court on consideration of the evidence on record came to the conclusion that the prosecution by circumstantial evidence has established the case against the accused, hence held the accused persons guilty and awarded capital punishment to A-3, the brother in law of the deceased, Suchindra Nath Mandal, while the other three accused persons were convicted under Section 302 read with 34 and were awarded life imprisonment.

Page after page of unstructured text

Legal Issues?

Arguments?

Reasoning by the court?

1. Original complainant in Sessions Case No.127 of 1989 before the Sessions Judge, West Dinajpur, at Balurghat is in appeal before us in the above Criminal Appeal, against the judgment of acquittal made by the High Court Calcutta in Criminal Death Reference No.4/1990 and Criminal Appeal No.327 of 1990. The State has not preferred any appeal against the judgment of the High Court, but is a party respondent before us.

2. Prosecution case briefly stated is that one Suchitra, the daughter of the appellant herein was married to Nakul Chandra son of Sarat Chandra Mandal, accused No.1 before the Sessions Court. At the time of the marriage, there was an agreement to pay dowry in cash of Rs. 500/- out of which, Rs. 300 was paid to the accused A-1 by the appellant but he could not pay the balance amount of dowry due to poverty. A-1 was living with his wife A-4 and five of his children of which A-3 Sachindra Nath Mandal is one of them who was living with his wife Gauri Mandal, who was A-4 before the Sessions Court. The husband of deceased Suchitra was Nakul Mandal, who was also staying with his father. It is the prosecution case that because of the non-payment of balance of dowry, there was constant torture and ill-treatment of the deceased by the accused persons. This was made known to the appellant (PW-1) as also his wife PW-1 and Sarat

- **Facts**

- **Ruling by Lower Court**

3. On 16.8.1986, the deceased Suchitra, daughter of accused No.1 (father in-law) and consumed poison in the house of accused No.1 (father in-law) and in view of the fact PW-11, husband of PW-1 was along with PW-2 away in the field, had proceeded towards the village of her sambandhi and on the way they met one Nakul Bhunia, who took them on his cycle towards Durlavpur. Further on the way, they found the body of deceased Suchitra being carried on Plank tied to the same by A-3 Sachindra Mandal with the help of some people and on seeing them A-3 and others left the body there and went to some distance. When PW-1 and PW-2 went near the body, they found bleeding from the mouth and nose. PW-1 and PW-2 took the deceased to the hospital at Naliganj to Balurghat Hospital, where PW-15 Dr. Nath examined the deceased and declared her as brought dead. He also opined that the death was caused by throttling. On the basis of the information, received from the Dr. PW-15, the police of the Balurghat Police Station registered a case under Section 302 of the Indian Penal Code. Dr. D. S. Shal PW-16 on 17.8.1986. He opined that cause of death of the deceased was due to manual strangulation and was homicidal in nature. During the post-mortem, he noticed the bruise mark on the neck and both on right and left side of the wind pipe. On dissection, he found the hyoid bone fractured. After completion of the investigation, the police filed charge-sheet against the above-mentioned four accused persons.

- **Precedent citations**

4. The prosecution has primarily relied upon the evidence of PW-1, the mother, PW-2 aunt, PW-4 brother, PW-11, the father and another relative PW-2, Govind Sarkar to establish the fact that there was a dowry demand, pursuant to which the deceased was being tortured and the time of finding the body of the deceased was during the time of taking the deceased to the same. The prosecution also relied upon the evidence of the prosecution that there was a homicidal death by throttling.

- **Ratio of the decision**

5. A very peculiar defence was taken by the accused persons who contended that on the fateful day, the deceased had consumed poison and when it was noticed by her husband, Nukul, he tried to prevent the deceased from swallowing the poison by pressing the neck at which time A-3, brother in law of the deceased allegedly came and put his finger into the throat of the deceased to make her vomit the poison. The further contention was that at that point of time, deceased was alive, therefore, they decided to take her to the hospital and it was during that time that the mother of the deceased met on the way when the body was being taken to the hospital. The Trial Court on consideration of the evidence on record came to the conclusion that the prosecution by circumstantial evidence has established the case against the accused, hence held the accused persons guilty and awarded capital punishment to A-3, the brother in law of the deceased, Suchindra Nath Mandal, while the other three accused persons were convicted under Section 302 read with 34 and were awarded life imprisonment.

A well-studied problem in Legal NLP:

Rhetorical segmentation / rhetorical role labeling of case judgements

Label each sentence/paragraph based on its rhetorical functionality

Challenges in NLP / IR over legal documents

Length of documents

- Very long documents
 - Tens to hundreds of pages
 - Avg #words in Indian Supreme Court case judgements: 4.3K
 - Avg #words in UK Supreme Court case judgements: 14.3K
- Problem in applying neural models
 - Legal docs often much longer than news articles, research papers, ... over which neural models have been designed
 - Most recent neural models cannot take a full legal document as input

Complexity of text

- Lengthy and complex sentences
- Non-standard ways of writing, frequent use of abbreviations
- Even identifying the end of sentences is challenging
 - [Sentence boundary detection in legal text, Natural Legal Language Processing Workshop 2019]

Complexity of text

- Lengthy and complex sentences
- Non-standard ways of writing, frequent use of abbreviations
- Even identifying the end of sentences is challenging
 - [Sentence boundary detection in legal text, Natural Legal Language Processing Workshop 2019]

This Court in Indian Petrochemicals Corporation **Ltd. & Anr. vs.** Shramik Sena & **Ors.**, 1999(6), referred to the decisions in Parimal Chandra Rahas case, Reserve Bank of India **vs.** Workmen, 1996, and **M.M.R. Khan vs.** Union of India, 1990, and held that the workmen of a statutory canteen, as in the present case, would be workmen of an establishment for the purposes of the Act only and not for other purposes.

[71 words]

Standard NLP tools do not usually work well on legal text

- E.g., standard Named Entity Recognizers (Stanford / Spacy) often incorrect

“Life Insurance Corporation, India”
Life : PER, Insurance Corporation : ORG, India : LOC

Example using
Stanford NER

On the basis of the said sale certificate the mutation Entry No.1836 was effected in the village record in favour of Gulabai Desai PERSON , and thereby her name was entered in Survey Nos.118/1B and 328 of village Lonavala PERSON to the extent of 29.30 Ares and 70 Ares respectively. Thereafter, Gulabai ORG sold CTS ORG No.133 admeasuring 33 Gunthas on 24.4.1977 DATE to Respondent No.3 Genu Kadu PERSON . The said Gulabai ORG also gifted her remaining area from this Survey numbers to her grandson Anil Gajanan Desai PERSON on 15.1.1979 DATE , who in turn has sold his properties to Respondent no.2 Prem Hasmatraj Lalwani PERSON in the year 1980 DATE .

Example using
Spacy NER

missed

Wide variation in style of writing

- Writing style and document structure varies widely from one country to another
- Lack of standard notations / terminologies for writing legal text
- E.g., summarization algorithms designed for documents of one country do not work well on documents from other countries [Bhattacharya et al. ECIR 2019]

Expensive to build large-scale datasets

- Difficult to create large datasets required for supervised NLP/IR methods
- Requires annotators having expertise in Law → expensive, time-consuming

Challenges → Opportunities

- Law practitioners' needs
 - Huge amounts of raw legal data made available digitally in the last 5-6 years
 - Need automated systems to make sense of all this data
- Common people's needs
 - Severe lack of knowledge of Law among the common masses
 - Expensive to consult Law experts in most countries
- Economic benefits of developing AI systems that are **easy to use for Law practitioners and the common man** and provide usable results → startups