

刑事裁判例からの量刑関係情報抽出の試み

野田 奏^{†1} 竹内 和広^{†1} 三島 聡^{†2}

過去数年分の刑事裁判の判決書を収集し、量刑に関わる情報の抽出・検索を目的とした、構造化文章データベースの構築を行っている。本稿では、左記データベースを構築する上での基本的な技術となる情報抽出用の言語表現パターンを自動的に構築する試みについて紹介する。また、それらのパターンを利用した、量刑情報キーワードの自動抽出についても検討する。

A Pilot Study on Retrieving Sentencing Information from Previous Judicial Decisions

SO NODA,^{†1} KAZUHIRO TAKEUCHI^{†1}
and SATOSHI MISHIMA^{†2}

In order to develop a judicial precedent retrieval system, we have to apply the NLP technologies to legal documents. In this paper, we propose an unsupervised method to automatically retrieve words and phrases that are relevant to the sentencing information from criminal court decisions delivered in the past few years. As an evaluation of the phrase-based analysis on the data, we confirm its efficiency by a keyword extraction task.

1. はじめに

2009年から、刑事裁判において裁判員裁判が行われるようになった。裁判員制度では、選挙人名簿をもとに無作為に選出された一般市民（裁判員）が、裁判官とともに重大な事件を

担当し、公判廷に出席して有罪・無罪を判断し、そして有罪の場合には刑の内容に関する量刑判断も行う。従来の職業裁判官による刑事裁判では、量刑判断は、過去の裁判例を踏まえつつ、裁判官自身の裁判経験を通じて得た知識をもとにして行われてきた。しかし、裁判員制度のもとで裁判員として刑事裁判に参加する市民には、このような知識が共有されていない。このことが、真の意味での市民の司法参加の実現にあたって問題とされている。実際、裁判員制度開始前に実施された裁判官及び市民向けの大規模なアンケート調査及び模擬裁判においては、個別の量刑事情が刑軽減事情なのか、刑加重事情なのかという評価方向、それがどの程度量刑を左右するのかという重要度について、裁判官は比較的統一的な見解を有していたのに対し、市民には同様の考え方の共有はみられず、評価方向・重要度ともにかなりばらつきがあることが示されている¹⁾。

また、裁判員裁判では、裁判員の主体的判断が尊重されるべきとされているが、以下のような量刑判断に関する問題が指摘されている。

- 裁判員ひとりひとりの感覚的判断に基づき、従来と大きく異なる形で量刑事情の評価がなされることになれば、量刑判断の予測可能性・公平性を害し、その妥当性に疑問が生じてしまう。
- 他方、裁判官の専門家としての知識をもとにして量刑判断をすれば、裁判員は裁判官に依存せざるを得ず、裁判官との対等性や裁判への主体的関与は確保できなくなる。
- そこで、少なくとも、裁判官の専門家としての知識の内容やその合理性が、裁判員にもわかるように、かみくだいて提示されることが必要になるが、量刑の審理や判断の方法にはあいまいな点が多く、法律学の分野でも詰められているとはいいがたい状況にある。

これまでの裁判員裁判の運用では、基本的に、従来の量刑相場から大きくずれない範囲で量刑判断がなされているようである。我々は、そのような審理や判断の参考情報として、既存の裁判例を複雑な条件下で多角的に参照する技術を提供することは有益であると考え、本稿では、裁判員裁判において適正・妥当な量刑審理・量刑判断を確保するための基礎資料を提供することを目的とし、既存の刑事裁判例の蓄積を計量的に分析するための基礎技術として、自然言語で記述された量刑に関する情報を抽出・蓄積し、それらを検索する技術について検討する。

2. 刑事判決書の検索システム

情報工学及び情報科学の分野でも、裁判の判決書をデータベース化する試みがなされており、自然言語で記述された判決書を検索するための、さまざまな工夫が提案されている。

^{†1} 大阪電気通信大学 大学院 工学研究科

Osaka Electro-Communication University, Graduate School of Engineering

^{†2} 大阪市立大学 大学院 法学研究科

Osaka City University, Graduate School of Law

例えば、江越らは、判例の多くが内容の異なる複数の「段落」からなり、各段落が階層構造を持つことに注目し、「原告の主張」「被告の主張」「裁判官の判断」部分に出現する法律判断用語の頻度に差があることから TF*IDF 法を用いて文書ベクトルに対する重み付けを行い、類似判例を効率良く検索する事に成功している²⁾。この結果は、文書検索において単純かつ一般的なキーワード検索の重みづけ法である TF*IDF 法の単純な適用が、判決書検索のような目的志向の強い専門文書検索においては必ずしも有効ではなく、専門文書の特殊性や特徴を生かしたデータベース化が必要なことを示している。

また、本多は、本研究と同様な目的意識から、裁判員制度において裁判員が過去の判例を検索することを念頭においた研究を行っている³⁾。具体的には、判例・法令検索データベースである、LEGALBase⁴⁾の約 3800 件に対して既存の概念ベースシステムを利用した判例文書検索システムを構築し、被験者の作成した検索クエリを評価に用いて、複数の手法によって検索システムの精度向上を目指した概念ベースのチューニングの検討を行っている。その結果、不要語の設定によるチューニングでは、品詞情報を用いた不要語の設定で検索性能の若干の向上が見られたのに対し、法律用語辞書知識の追加では、辞書知識の重みを変化させて検討したにもかかわらず、検索性能の向上が見られなかったとしている。複数のチューニング手法のうち、最も効果があったと報告されているのは、文書集合をあらかじめ、罪状ごとのクラスタに分類しておき、検索対象を絞る方法である。また、課題として、判例文書がもつ文体の特殊性の問題を指摘している。

以上の関連研究を踏まえて、本稿でも判決書の構成や文体の検討がまず必要と考え、本稿が対象とする、刑事判決書の一般的な意義・特徴を整理すると以下ようになる。

- 刑事事件において裁判長が言い渡した判決の内容を証明する文書
- 担当した裁判官が作成する。
- おおむね、次のような内容が記載される。
 - － 事件番号、被告人を特定する事実（本籍・住所・氏名・生年月日）、主文（「被告人を懲役〇〇年に処する。」など）、主文の理由
 - － その理由の箇所では、有罪判決にあつては、有罪と認定された事実、有罪の基礎になった証拠の標目、法令の適用、弁護人などの主張に対する判断、量刑の理由などが記される。
 - － ただし、本研究に利用した最高裁判所ウェブサイトに掲載されている判決書には、事件番号、被告人を特定する事実の記載がなく、また、証拠の標目や法令の適用も原則として「省略」とされている。

- 判決書ごとに何がどこに書かれるかという文書構造にばらつきがある。また、言語表現も統一されていない。

以上のような特徴をもつ判決書の記述のうち、本稿では特に、量刑に関する情報を主眼にデータベース化することに焦点を絞る。この方向性は、先に挙げた先行研究と対立するものではなく、判決書文に対して量刑に関わる条件付けによる検索方法を提供することを主眼に据えるにすぎない。具体的には、判決書に記述された個々の量刑の理由について、以下のような分類がなされるがそれぞれの記述は自然言語によるものであるため、判決書間の類型を計量的に分析することが困難である。そこで、本研究では、以下のような事実に関する言語表現を抽出し、それぞれの分類項目を観点とした検索への対応を目標とした、判決書文のデータベース構築と検索方法について検討する。

- 犯情に関わる事実
犯罪事実と直接関係する事実。具体的には犯罪の種類・罪質、犯行態様、犯行の結果、犯行の動機・目的・計画性、共犯との関係・関与の程度など。
- 一般予防に関わる事実
国民一般に対する犯罪の予防に関係する事実。例えば、模倣性が高いので重い処罰を科すという場合の高い模倣性を基礎づける事実。
- 特別予防に関わる事実
被告人本人による再犯の防止に関係する事実。例えば、本人の反省、監督者の存在、薬物事案における薬物離脱のプログラムや自助グループへの参加など。
- その他の政策的考慮に関わる事実
上記に含まれない事実。例えば、当該事件により被告人が社会的制裁を受けた事実、捜査機関により違法な捜査が行われた事実など。

3. 量刑理由に関する文書構造分析

3.1 調査対象としたデータ

前節ですでに紹介した江越らの論文では、「判例は一つの文書ではなく、役割の異なる複数の文書に分けることができる」という知見から、判決書の文書構造を利用して判決書のデータベース化を図り、高度な検索を可能にする工夫を行っている。本稿でも同様に、判決書の量刑に関する情報記述の文書構造の調査を行った。

具体的には、最高裁判所ウェブサイト⁵⁾の判例検索システムにおいて、平成 16 年以降、地方裁判所で審理されたもののうち、「殺人」というキーワードを含む判決書を 303 例収集

した。なお、判決書はPDF形式で提供されているため、これらを計算機で容易に処理できるようテキスト形式に変換し、今回の検討に利用する。ここで、地方裁判所で審理され、かつ、「殺人」というキーワードにより調査対象の文書を絞り込んだ理由は、裁判員制度の対象となる重大事件であること、それら事案において、量刑に関する情報記述・表現にどの程度の多様性があるかを検討するためである。

上記303例の調査の結果、量刑に関する記述は「量刑の理由」などの表現を使い、判決書の一部としてまとめて記述されることが多く、その記述は構造的な文章となされることが多い。他方、文の長さや言い回しといった言語表現は、判決書ごとに多様性が高いという知見も得た。以上の知見より、量刑に関する情報の記述の類型化が必要であると考えた。

3.2 文章構造分析

上記の条件で収集した判決書のうち、「量刑の理由」などの明確な記述を伴い、部分文章として量刑に関わる情報が言語によって記述されている文章の構造分析を行った。このような文章は、冒頭の数文に事件とその裁判に関する概要説明がなされ、また、文章末には量刑を結論付ける帰結のための数文が記述される大まかな構造をもっている。しかし、冒頭の概要説明と文章末の帰結の数文以外は、2節で述べた、「犯情に関わる事実」「一般予防に関わる事実」「特別予防に関わる事実」「その他の政策的考慮に関わる事実」が当該事案の具体的な説明とともに記述されており、そこで使われる表現も多様性が高い。例えば、犯行態様、犯行の結果、犯行の動機・目的・計画性などの記述がなされている判決書は多いが、その記述方法は多様であり、記述順序も不定である。また、言語表現が文、句、文節、語といった言語単位のどの単位で類型化が図れるか（あるいは類型化できないか）については、直観的な判断は難しい。さらに、すべての判決書に「共犯との関係・関与の程度」や「高い模倣性を基礎づける事実」、「捜査機関により違法な捜査が行われた事実」といった記述がなされているわけではない。

他方、文章を構成する上での文の機能的な特徴はみられる。文長には多様性があるものの、上記の事実に対して、評価判断を明確に下している文が存在する点である。例えば、以下のような文である。

- 「...暴力で排除しようという凶暴で身勝手な考え方が、極めて厳しい非難に値する。」
- 「...経緯及び自己中心のかつ浅はかな犯行動機に酌量の余地はない。」
- 「...被害感情が厳しいのも無理からぬところである。」

このような、明確な評価判断の記述を基準に、判決書の構造化や類型化が可能かを念頭に置きつつ、形式的な言語記述から客観的な判断が可能か文の機能類型を試みた。その知見か

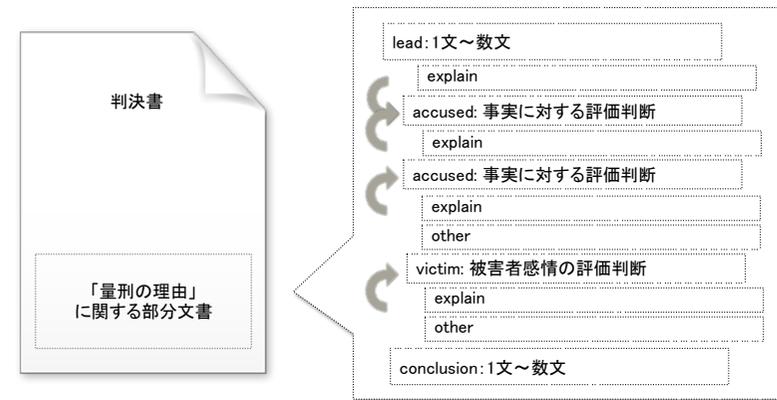


図1 判例の構造
Fig.1 text structure analysis

ら実験的なタグ体系を構築し、量刑理由を記述した部分文書の構造化を行った。タグ体系を表1に、これによって分析した判決文書の構造イメージを図1に示す。表1の体系は、実験的なものであり、今後、階層化や詳細化が可能である。

また、この量刑理由部分を構造分析する過程で、量刑理由の類型化可能な要素には、それぞれ特有の表現方法があるという直観が得られた。この直観が正しければ、文の表現上の特徴によって量刑情報の記述類型ごとに文の分類やそれに関わるキーワードの抽出を自動化し、量刑に関する情報を精細に分析できる可能性があると考えた。そこで、本稿の残りでは、この可能性について、計量的な方法により検討する。

表1 量刑理由の構造化のためのタグ一覧
Table 1 Our annotation tags for text structure analysis

タグ	意味
lead	事件とその裁判に関する概要説明
accused	被告人の行動などを明確に非難・評価した記述
victim	被害者や被害者家族の事情・心情推察に関する評価記述
conclusion	結論
explain	前後の評価記述文に対しての背景/補足情報の記述や説明
other	上記以外の記述

4. 判決書からの構文パターンの抽出

本節では、日本語の文の表現上の特徴を機械的に計量できるようにするために、文節単位での係り受け関係を利用して、構文的特徴を持った表現の型（以降構文パターンと呼ぶ）を抽出することを試みる。これは、文の内容的な語を空項にして機能的な語と組み合わせた表現の型であり、文のもつ機能的な役割のみを特定するものである。

入力された文に対し、以下の処理を順に施すことによって、構文パターンを生成する。

- (1) 判決書中には記号を伴った箇条書きや特殊な記法によって記された事件番号、裁判の日付などが存在し、これらが以降の解析時にノイズとなるため、テキストから取り除くか、代替記号に置換するクリーニングと呼ばれる作業を行う。また、ここでは日本語の文のみを解析の対象としたいため、空白、句点及び改行でテキストを区切り、1文中に5%以上のひらがなが含まれる5文字以上の文のみを抜き出す。さらに、括弧書きされた文は以降の係り受け解析時にうまく処理することができないため、正規表現を利用して取り除く。
- (2) 日本語係り受け解析器 CaboCha⁶⁾ を使用して文節ごとの係り受け関係を解析し、構文木と呼ばれる、文節をノードとする木構造データを獲得する。このデータ構造では、あるノードが複数の子ノードを持つ場合、それ以前の文節からの複数の係り受け関係を持つことを表している。
- (3) 各ノードの中で、内容的な意味を持つサ変名詞・サ変動詞を <S>、単名詞・複合名詞を <N>、サ変以外の動詞を <V> のようにそれぞれ品詞を表す記号に置換する。日本語の機能的な役割を持つ接続詞、副詞、助詞、助動詞などは置換しない。また、出現頻度の高い名詞150個程度と形容詞については構文的特徴を限定すると考えられるため、これらも置換しない。
- (4) 最後に木構造の根部分から各葉へ順に n 個の要素を辿り、元の順に並べた文字列を構文パターンとして抽出する。

以上の一連の処理の流れで、例として「そして、犯行の動機は誠に身勝手なものであり、犯行に至る経緯にも酌むべき事情は認め難く、被告人の犯罪性向は誠に顕著である。」という入力文に対し、 $n = 4$ としたときの解析の様子を図2に示す。この例での出力は「<N>な<N>であり<N>は<N>である」と「<N>は<N>難く<N>は<N>である」の2つの構文パターンとなる。

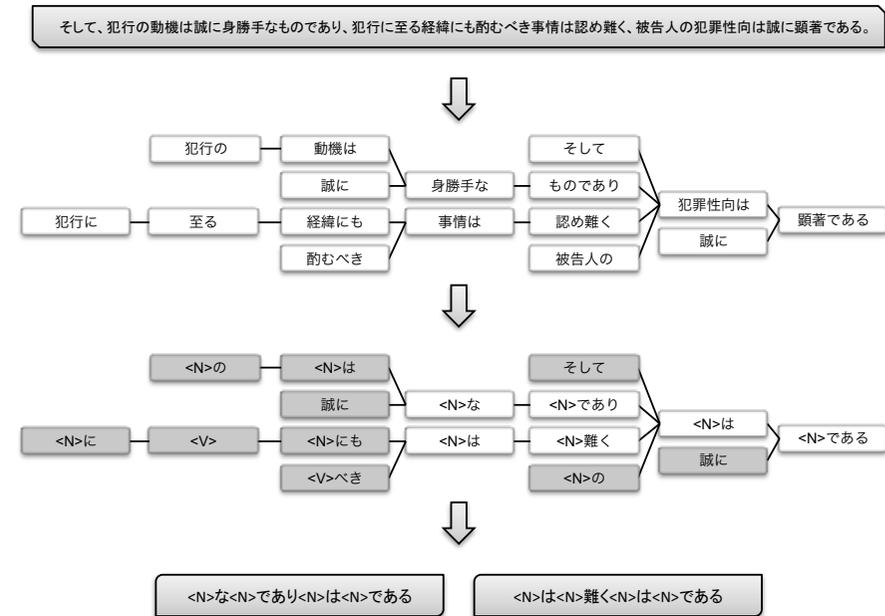


図2 係り受けを考慮した構文パターンの生成

Fig.2 An example of generating of construction patterns from the dependency structure

5. 量刑情報キーワードの抽出

収集した303例の判決書の量刑理由部分に対して、前節で述べた生成手法を用いて構文パターンを生成した。生成した構文パターンと、キーワードの例を図3に示す。

生成した構文パターンのうち、その構文パターンに当てはまる文が少ないものについては独自性が高く、構文パターンとしての重要度はそれほど高くはないと考えられる。そこで、図4に、 $n = 2$ から6に変化させたときに、全量刑理由文のうち、2つ以上の量刑理由文に構文パターンが合致する割合と、構文パターンとただ1つの量刑理由文しか対応しない場合の文の割合を示す。図3及び図4より、 $n = 4$ 程度の構文パターンが量刑事情文を解析するのに有効だろうという感覚を得た。

ここで、「量刑の理由」の文章構造分析と、 $n = 4$ の文節を持つ構文パターンの分布を確認

Pattern	<V>た	<N>の	<N>は	<N>である
	得た	財政上の	額は	多額である
	失った	遺族らの	処罰感情は	熾烈である
	果たした	犯行の	結果は	甚大である
Pattern	<N>を	<V>	<N>は	<V>ない
	態度を	窺う	ことは	できない
	責任を	負う	理由は	見当たらない
	非難を	受ける	ことは	免れない
Pattern	<N>に	<V>た	<N>は	<N>である
	周辺地域に	与えた	不安感	深刻である
	各共犯者に	及ぼした	影響力は	重大である
	犯行に	果たした	役割は	多大である

図3 生成した構文パターンと獲得したキーワードの例
Fig.3 Example of generated construction patterns and extracted keywords

する。n = 4 の構文パターンが2回以上マッチする文は図4のように40%程度である。また、文章構造分析の各タグにおける詳細な分布を示したグラフを図5に示す。図5のように、タグがつけられた評価判断機能をもつ文には、機能語の定形表現である構文パターンが多く分布することが見て取れる。すなわち、多様性の高い量刑理由の記述においても、評価判断はある程度一定の形式で記述がなされていると考えられる。そこで、このことを利用して、先に示した構文パターンとその空項を埋める内容語を抽出した結果を検討する。さらに、比較対象として全ての量刑理由文から単／複合名詞を単純に抽出する。

構文パターンを利用した内容語抽出にはaccusedタグをもつ文から抽出した構文パターンを利用した。それぞれの方法で抽出した語の中で、頻度が高いものから順に20語ずつを表2に示す。

表2の中で両手法で抽出した語の中で同じように上位に出てきた語は「犯行」だけである。さらに、構文パターンによって抽出した語には「執拗」、「残忍」、「悪質」、「非情」、「冷酷」といった、accusedタグを付与した文に多い、被告人を非難するのに時に使われる語が抽出できていることが見てとれる。つまり、構文パターンによって抽出される語が一般的な

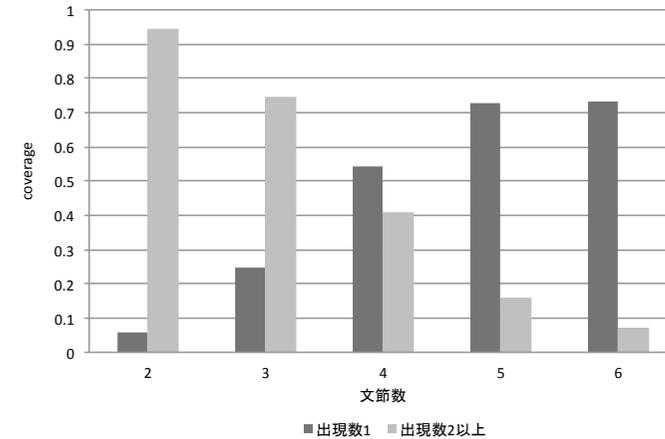


図4 構文パターンによる量刑理由文の網羅性
Fig.4 The cover ratio of phrase patterns

語とは広くかけ離れており、より記述内容に則した語、つまりは判例検索などにおいて文書の特徴付けるのに有効であると思われる内容語が抽出できていると言える。

6. おわりに

本論文では、刑事裁判の判決書の量刑情報に関わる記述に対して、高度な分析・検索に資

表2 構文パターンと単純な手法によって抽出した内容語

Table 2 extracted keywords

構文パターンを用いたもの		単純に抽出したもの	
犯行	執拗	被告人	事案
重大	被害者	被害者	自ら
悪質	強固	犯行	動機
被告人	非道	本件	死体
刑事責任	殺意	事情	態度
残忍	自己中心的	本件犯行	自己
冷酷	身勝手	重大	悪質
非情	本件	事件	遺族
態様	刑責	結果	旨
結果	短絡的	同人	共犯者

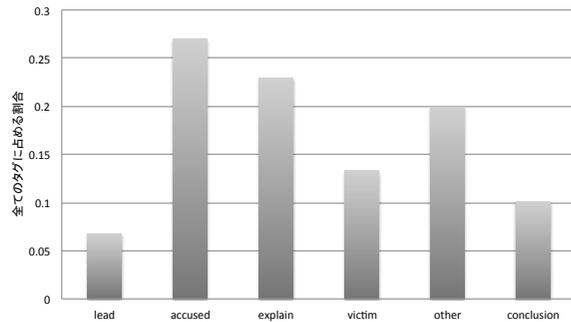


図5 構文パターンが合致する文タイプの分類

Fig.5 The distribution of sentence types in the pattern-matched sentences

する技術を、文章の構造分析の調査と構文パターンによる検索キーワードの絞り込みの両面から検討を行った。その結果、判決書の量刑情報に関わる部分には評価判断に関わる機能的な役割を果たす文が存在し、そのような文には本研究により抽出した構文パターンに合致する記述が相対的に多く使われていることを確認した。また、そのような構文パターンを用いて、量刑情報に強く関わるキーワードの抽出を行った。これら抽出したキーワードが、出現頻度に基づく単純な手法によって抽出されたキーワードに代わる、文の機能的な性質に即した役割をもつのかを今後さらに検討していきたい。

謝辞 本研究は、財団法人日本証券奨学財団の2010年度研究調査助成金「裁判員裁判において当事者の主張提示が量刑および判決に与える影響の検討」(研究代表・三島聡)による研究成果の一部である。同財団には謝意を表したい。

参 考 文 献

- 1) 司法研修所編：量刑に関する国民と裁判官の意識についての研究—殺人罪の事案を素材として—, 法曹会 (2007).
- 2) 江越裕紀, 片上大輔, 新田克己：判例の構造を利用した判例文書検索, 情報処理学会研究報告, Vol.2005, No.11, pp.1-8 (2005-01-28).
- 3) 本多弘幸：裁判員のための判例検索システムの概念検索性能の向上, 情報処理学会研究報告. 自然言語処理研究会報告, Vol.2009, No.2, pp.1-6 (2009-11-09).

- 4) 日本法律情報センター：LEGALBase 判例 法令 検索サイト, 株式会社日本法律情報センター (オンライン), 入手先(<https://hanrei.lawsquare.jp/doc/index.php>) (参照 2011-07-04).
- 5) 最高裁判所：裁判所ウェブサイト, 最高裁判所 (オンライン), 入手先(<http://www.courts.go.jp/>) (参照 2011-07-04).
- 6) 工藤 拓：CaboCha: Yet Another Japanese Dependency Structure Analyzer, (online), available from (<http://chasen.org/~taku/software/cabocha/>) (accessed 2011-07-04).