

## コーパスを活用した法文データの分析に関する問題点

矢野 信 (株式会社法学館法教育研究所) †

### Problems in Corpus-based Studies of Legal Codes

Makoto Yano (Japan Research Institute of Law Related Education, HOUGAKUKAN CO., LTD.)

#### 1. はじめに<sup>1</sup>

『現代日本語書き言葉均衡コーパス』(BCCWJ)の「特定目的サブコーパス」の一つに「法律」がある。これは、「公的な性格の強い書き言葉であり、これらの分析により言語政策に関わる基礎資料を提供することが期待できる」(丸山(2011))として、独立したサブコーパス(以下「SC」という。)とされている。

しかし、例えば第1~3回の「コーパス日本語学ワークショップ」において法律SCに言及する発表は11件であるなど、その分析・研究がまだまだ活発であるとはいえない。

本発表では、コーパスの手法によって法文を分析する前提として、①法律(法文)の言語を分析することの位置付け、②法律SCやその他のデータベースのコーパスとしての性質について、法律実務家・法律教育者側の視点を交えながら検討する。加えて、末尾で法律SCなどを用いて得られるデータの例を挙げる。

#### 2. 法律(法文)の言語に関係する先行研究

松田(2011)などでは、法令の文言を純粋に言語資料として用い、法令の言葉の中に見られる「ゆれ」の観察から、変異の内的要因についての考察を行なっている。そこでは、法言語学との関係についても言及がなされている。これは、分析・考察の方向性や関心において本稿とは異なるが、手段の第一として法令の言語そのものを分析する(そこにコーパスの手法を用いる)という点で共通性を有する。

このほか、法律の言葉について、法哲学者・言語哲学者による研究は数多くあるが、言語学のアプローチをとるものではないことから、ここでは省略する。法的推論を計算機で実現するという研究(自然言語処理・人工知能の観点)についても同様である。

#### 3. 法律(法文)の言語を分析することの位置付け

まず、本稿で用いる「法文」という概念について定義をした上で、(コーパスの手法による、よらないに関わらず)「法文」の言葉を分析することの位置付けを確認しておく。

##### 3.1 用語の定義

###### 3.1.1 「法律」

これは、大きく分けて、①国会の議決を経て成立する法形式としての「法律」の意味、②法律(法)に関係する分野・領域の2つの意味で用いられる(金子(2008))。<sup>2</sup>

† klagegrund@gmail.com

<sup>1</sup> 法律の条文は、本来、縦書き・漢数字で表記されているが、本稿では、横書き・算用数字で表す。また、条文中の読点は本来「、」であるが、本稿では「,」に置き換えた。

なお、根拠として法令の条文を挙げる場合、法令名+条文番号のみを掲げ、書物あるいはインターネット上の法令集を文献として掲げることはしない(ただ、インターネット上でアクセス可能な法令集を末尾の関連URLに記載しておく。)

<sup>2</sup> 一般に“法律の言葉は難しい”といわれるときは、①②どちらの意味で使われているか分からない。

本稿では、「法律」の語を基本的には①の意味でしか用いていない。②の意味では、「法律分野」などとする。

### 3.1.2 「法令」

公権力によって制定され、成文の形式をとる規範のことをいう（条約は含まない）。これには、上記の意味の「法律」のほか、内閣が定める「政令」、各省庁の「〇〇省令」、地方自治体の「条例」、衆議院・参議院の議院規則、裁判所が定める裁判所規則などがある。（一般的な定義、法学上の定義と同じ意味で用いる。）

### 3.1.3 「条文」

法令を構成する個々の文のことを指す。法令を構成している規範（ルール）の最小単位であるといえる。“項目”という意味合いを込めて「条項」ということもある。次の「法文」と区別して使う。

### 3.1.4 「法文」

法律・法令の内容を構成している文のことをいうが、この言葉は、法学書・法律学辞典などで定義がなされていない。<sup>3</sup> 本稿で今後、分析の対象として取り上げるのは、この意味の「法文」である。<sup>4</sup>

従来、この語は、主として法哲学・言語哲学の観点から、規範（ルール）の内容を表している言葉の意味で用いられることが多かった（規範が言葉によって表されるということについての関心である。）。

ただ、本稿では、言語資料（書き言葉）としての法律・法令を分析するという観点から、成文化された法令の内容を構成する言葉を含むことにする。その中心的部分は、従来の意味の“法文”であるが、それだけではなく、例えば制定文<sup>5</sup>、前文<sup>6</sup>などの部分も含む。

## 3.2 「法文」と類似の性質を有するものの例

上記のような意味の「法文」の言葉は、コミュニケーションのために用いられるものではなく、個々の法文の言葉を発するという発話行為も観念できない。<sup>7</sup> このような意味の「法文」には、「法令」を構成する言葉のほか、表 1 に掲げるものがみな含まれると考えられる。

表 1 「法文」と類似の性質を有するものの例 (◎◎の区別は後に触れる)

- |   |  |
|---|--|
| ◎ | 解釈通達といわれるものの一部（所得税基本通達など。法令の解釈に関して、条文の形で作られている。） |
| ◎ | 「行政指導指針・指導要綱」といわれるものの一部                          |
| ○ | 各種組織の内部規則（例：〇〇大学学生委員会規則）                         |
| ○ | 「公認野球規則」, 「旅客営業規則」, 就業規則, 約款, 定款, 町内会規則, 校則      |

<sup>3</sup> 末川(1991), 金子(2008)などには項目が存在しない。また、町田(1980)でも特に定義をせずに用いている。

<sup>4</sup> なお、Tiersma (アメリカの法言語学者) は、文書化された法の言葉を、①効力のある文書（法律・判決文・契約書など）、②解説文書（依頼者へ事件を説明する文書、法律の教材など）、③説得を目的とした文書（裁判所に提出される準備書面など）の3つに分類しており（Gibbons(2013)）、ここでは法律と判決文とが同分類とされている。しかし、筆者の当面の目的として、規範それ自体を表した言葉とそれ以外との分離を行いたいことから、さしあたり、その分類にはよらないこととした。

<sup>5</sup> 法令の冒頭に置かれ、その法令の制定根拠を表現するための文章のこと。例：「弁理士法（大正10年法律第100号）の全部を改正する。」（弁理士法（平成12年法律第49号）の制定文）

<sup>6</sup> 法令の冒頭に置かれ、その法令の制定趣旨・目的などを述べた文章のこと。例：日本国憲法前文

<sup>7</sup> なお、いわゆる「公布文」（法令の公布者の意思を表明する文章。「〇〇法を公布する。」）は、行為遂行的な言葉（発話行為）である。

また、対等な当事者間で作成される「契約書」「和解調書」などの文書を構成する言葉も、上記のような意味の「法文」の言語的性質を備えていると考えられるが、その点自体検討を要する問題であることから、ひとまずは分析の対象から外しておく。

### 3.3 「法文」の分析の意義

本稿では、「法文」の言葉を分析する意義について、法令を“使う”側の視点（法律実務家・法律教育者側の視点）から、次のような点を意識しておきたい。

いわゆる「法律」の読み方については、法学部などの授業で最低限の事項は触れられるものの、法文の作り方について体系だって整理されているものは、いわゆる「法制執務」の実務の現場にしかほぼ存在しない。<sup>8</sup>しかし、前項のとおり、「法文」と類似の言語的性質を有するものには多様なものが含まれ、日常生活でごく一般的に触れるものもある。

「ルール」が広く集団社会一般に必須のものであるとすれば、法律家等の専門家教育の場面にとどまらず、法文の読み方・作り方についてのリテラシーを向上させる必要がある。「法文」の言語の分析をすることで、そのようなリテラシー向上に資する知見を提供することが考えられる。

以下では、そのような「法文」全体の分析のための第一歩として、主に「法律」を分析することを考えていく。

## 4. 法律 SC やその他のデータベースのコーパスとしての性質

### 4.1 「法文」を言語的に分析する際の留意点

#### 4.1.1 「法令」の著作権

著作権法 13 条には、権利（著作権）の目的とならない著作物として、「憲法その他の法令」（1号）、「国若しくは地方公共団体の機関（中略）が発する告示、訓令、通達その他これらに類するもの」（2号）などが定められている。

表 1 で◎としたものはみな 2 号に含まれる。一方、○としたものは、制定主体が独立行政法人などの場合には 2 号に含まれる。

これら著作権の目的とならないものは、コーパスを公開する際に著作権の処理を必要としない点がメリットとして指摘されている（松田(2011)）。

#### 4.1.2 法令が通用する時点・時間の問題

法令の言葉が他の種類の書き言葉と異なる最大の性質として、それが通用している時点・時間の問題がある。

法令以外の書き言葉はみな、それが作成された時点における言語という、いわば歴史の意味しか持たないと考えられるのに対して、法令の言葉は、それが効力を有するものとして通用している限り、“現在”の書き言葉としての側面を有し続ける。つまり、過去に制定・公布された法令がみな、現在の我々の日常生活を直接規律する。<sup>9</sup>

法令を“使う”側の視点においては、現在（現時点）において有効なすべての法令が関心の対象となり、その点を踏まえてコーパスのデータを分析・考察する必要があるろう。

#### 4.1.3 いわゆる改正法令の問題

前項に述べた法令の通用時間の点に関連して、いわゆる改正法令をどのように扱うかと

<sup>8</sup> 公共政策大学院では、法制執務に関する講義が開講されているところもある。

<sup>9</sup> ちなみに、「法令データ提供システム」に収録されている最古の法令は、（法律ではないが）明治 5 年太政官布告第 337 号（改暦の布告）である（1872 年）。これは、大日本帝国憲法の公布（1889（明治 22）年）より 17 年も前のものであり、これが、現在も有効な法令として存在する。

いう問題ある。改正法令とは、ある法令を改正するために制定される法令のことである（以下、「法律」を例に説明するが、原理的には「法令」のすべて、また、表 1 に掲げたものについても当てはまることである。）。

既に施行されている法律 X を改正する際、法律 X を改正するための法律 Y が公布・施行されることによって、法律 X の内容が X' に修正・変更される。このとき、法律 Y は、法律 X に「溶け込む」ことによって“消えてしまう”。

「法令データ提供システム」においても、市販されている各種の法令集（六法）においても、（当然ながら）溶け込み処理がなされた後の X' の状態（に加えて、法律 Y の附則部分。制定後附則といわれる）が収録されている（図 1）。

国会が制定した歴史的文書の全体という意味では、X/Y の側を見るべきであるともいえるが<sup>10</sup>、法令を“使う”側の視点においては、主に X' の状態に関心がある。研究目的に応じて何をコーパスの母集団と見るかにおける留意点の一つであると考えられる。

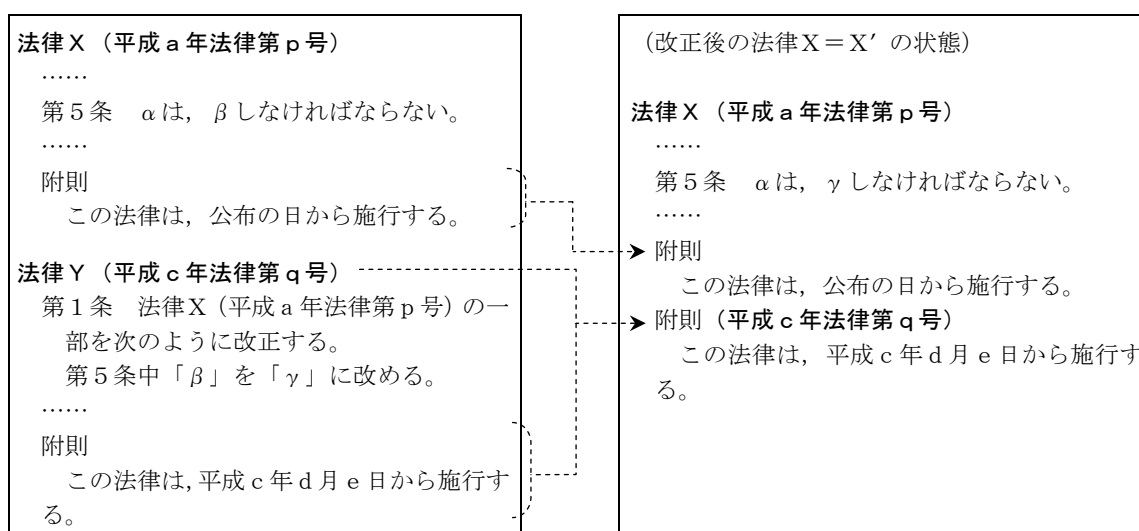


図 1 改正法令と溶け込みの関係

## 4.2 BCCWJ / 法律 SC のデータについて

ここでは、BCCWJ の法律 SC に収録されているデータの特徴について、丸山（2011）などから整理をした。

### 4.2.1 「法律」の原文との差異

データの真正性の問題として、法律の原文との違いの可能性を確認しておく必要がある。

法律 SC に収録されているデータは、「法令データ提供システム」に収録されている電子データをそのまま用いている。このことから、紙ベースの資料を電子データ化する場合に生じるような誤りの可能性は存在しない。<sup>11</sup>

ただ、「法令データ提供システム」に収録されている法令の電子データは、“正しい”ものであることが保障されてはいない。あくまで、法令として“正しい”ものとして効力を有するのは「官報」に掲載された状態であるというのが建て前である。<sup>12 13</sup>

<sup>10</sup> 「日本法令索引」では、このような改正法律を含む、全ての制定法令の全文データが検索可能である。

<sup>11</sup> HTML タグ除去の段階の問題は残る。

<sup>12</sup> この意味で、「法令データ提供システム」はあくまで、日本国の公式法令集ではない。

また、ブラウザでアクセス・検索可能な状態 (HTML ファイル) にする必要上、外字・ルビ・数式などの表示は原文とは異なる。また、システムの管理上、原文とは異なる処理をしている部分もある。

これらの点については、同システムの「最初にお読みください」<sup>14</sup>等に掲載されている注意点がそのまま当てはまる。

#### 4.2.2 収録範囲の点

データの代表性に関連して留意すべき点が、収録範囲の点である。

BCCWJ では、法律 SC も含めて、テキストの収録年代範囲が限られている。法律 SC については、「1976年から2005年までの間に公布され、2009年9月の時点でも施行されている法律」が「対象データ」であるとされている。

この対象データには、前記の意味の改正法律は含まれていないが、さらに細かく、次のようなことが見られた。

- ・ 制定時附則の中に含まれる「改め文」(他の法律を改正する条文)は、対象データに含まれている。<sup>15</sup>
- ・ いわゆる「整備法」(大規模な法律(例えば商法、民事訴訟法など)を全面改正する際に、多数に上る改正規定・経過規定<sup>16</sup>などをまとめて作られる単独の法律)については、対象データに含まれている。<sup>17</sup>

ただ、「整備法」の本則部分の「改め文」は含まれていないと考えられる。<sup>18</sup>

これは結局、BCCWJへ収録するために「法令データ提供システム」へアクセスした時点(2009年9月時点)において、「法律」として収録されていたデータのうち、廃止法令等<sup>19</sup>を除いたものと考えられる。上記の「対象データ」の意味は、このようなものとして理解する必要がある。<sup>20</sup>

#### 4.2.3 書誌情報データ「出版年」の意味

法律 SC のデータには、書誌情報「出版年」として各法律の公布年が付与されている。

ところが、前記の述べたような改正法令/溶け込みの仕組みによって、実際には、「出版年」よりも後の年に生成された言葉が混入していることがある。

例として、民事執行法(昭和54年法律第4号)は、「出版年/1979年」として表示されるが、このうち、「財産開示手続」に関する同法第196条以下の条文は、2004(平成16)年に施行された改正<sup>21</sup>によって追加されたものである。

検索結果として得られた各データが何年に生成された言葉であるかの点を厳密に確認す

<sup>13</sup> 本稿執筆中に発見したのものとして、「仮登記担保契約に関する法律」(昭和53年法律第78号)附則4条中の文言がある。「法令データ提供システム」「少納言」では、「ものである。」と出てくるが、正確には「ものである」(句点が不要)である(官報の画像データ(脚注22参照)で確認)。

なお、「日本法令索引」のデータは、正しく、「ものである」となっていた。

<sup>14</sup> <http://law.e-gov.go.jp/readme.html>

<sup>15</sup> 例えば、国際捜査共助等に関する法律(昭和55年法律第69号)の制定時附則の文言など多数。

<sup>16</sup> 旧法時代に起こった事態の効力などを定めるための条項(鈴木(2003)など参照)

<sup>17</sup> 実際、郵政民営化法等の施行に伴う関係法律の整備等に関する法律(平成17年法律第102号)の制定時附則の条文が「少納言」の検索結果に現れた。

<sup>18</sup> 法令データ提供システムに収録されないことから。

<sup>19</sup> 廃止法令、失効法令、実効性喪失法令のこと

<sup>20</sup> 例えば、動詞「改める」の頻度を得ようと法律 SC を検索するのは意味がないと考えられる。

<sup>21</sup> 「担保物権及び民事執行制度の改善のための民法等の一部を改正する法律」(平成15年法律第134号)による。

るには、その法律についての改正法律をすべてたどっていくしかない。<sup>22</sup>

#### 4.2.3 その他の点

法律 SC においては、「法令データ提供システム」に設定されている 50 の事項分類が書誌情報「ジャンル」として設定されているが、BCCWJ に収録する際のサンプリングの結果として、全部で 43 項目になっている。ジャンルごと／年代ごとの語数の一覧表<sup>23</sup>には、法律 SC の収録年代範囲を 5 年ごとに区切ったセグメントで、ジャンルごとの語数が掲載されているが、そこでは、全 258 個の小セグメント中 143 個 (55.4%) が語数 0 となっている。これらのことから、法律 SC を「ジャンル」「年代」を区切って検索した結果は、頻度を算出するのに意味を持たないことが多いと考えられる。

また、収録年代範囲が限られていることから、法律用語の検索には向かないことが多いと考えられる。前述のように法律は明治期以来のものが現在そのままの形で有効に存在し、かつ、改正を経ても原則として法律番号が変わらないことから<sup>24</sup>、我々の日常生活に密接に関連する基本的な法律<sup>25</sup>は、法律 SC にはあまり含まれていないことが予想される。

### 5. 法律 SC などを用いて得られるデータの例

#### 5.1 対照で用いたデータ

「法令データ提供システム」では、システムにアクセスした時期に有効な<sup>26</sup>現行法令のデータを全文入手すること可能であり、このデータ (以下「MIC データ」という。<sup>27</sup>) をもとにコーパスを構築することができる。「法令データ提供システム」のサイト上では、任意語による単純な全文検索は可能であるが、正規表現など複雑な検索には対応していないことから、複雑な分析を行うには、必要な範囲のデータをダウンロードして用いることになる。<sup>28</sup>

一方、BCCWJ の検索アプリケーション『中納言』では、形態論情報を用いた検索、共起条件を指定した検索などが可能であること、他のレジスターとの比較が可能であることなど、MIC データのみでは得られない情報を得られる可能性がある。

#### 5.2 検索結果の例<sup>29</sup>

今回、検索結果のサンプル例として、法律分野の言葉に特徴的であるといわれてきた 2

<sup>22</sup> 主要な法律については過去の法令集 (六法全書) を調べればよい。また、原理的には、「日本法令索引」の制定法令を検索することで可能である。政令・規則 (〇〇省令など) については、過去の官報を検索するしかない (官報には、全ての法令の原文・全文が掲載され、それによって法令が「公布」される。)。過去の官報は、「官報情報検索サービス」で全文検索が可能である (有料)。

<sup>23</sup> 「中納言オンラインマニュアル」に掲載されている。

<sup>24</sup> 例えば、「民法」は明治 29 年法律第 89 号、「刑法」は明治 40 年法律第 45 号といった具合である。

<sup>25</sup> 多くの人がイメージする法律としては、明治期に近代国家としての法整備をするために制定された法律群 (民法、商法、刑法など) や、戦後間もなくの時期に新憲法下での新しい国家体制を整備するために制定された法律群 (地方自治法、国家公務員法、労働組合法など) のものが多いと考えられる。これらは、実務でよく使われる法律でもある。

<sup>26</sup> システムの更新頻度の都合から、アクセス日現在の法令に厳密には一致しない。更新は約 1 か月ごとに行われているようである。

<sup>27</sup> MIC は総務省の英語略記である。

<sup>28</sup> 今回、MIC データをダウンロードしてテキストファイル化するところまでを自動化する Perl スクリプトを作成して、2013 年 5 月 20 日現在でアクセスして得られた MIC データを分析に使った。

<sup>29</sup> ここで示す MIC データは、法律 SC の収録対象データと同じく、法律番号が昭和 51 (1976) 年から平成 17 (2005) 年の間の法律のみから検索した結果である。ただ、元データの取得時点が異なることから (脚注 28)、法律 SC のデータと厳密には同じではない (法律 SC のデータ取得時点以降に改正・廃止された法律の部分が異なることになる。)

つの表現について、頻度を算出した。

### 5.2.1 「この限りで(は)ない。」

法文のただし書きなどで、一定の条件の下で本体のルールが及ばないことを表す意味に用いられる。検索結果を表 2 に示す。

- ・ 「は」が入っているかどうかで法文としての意味は変わらないと思われるが、実際に検索で出現するのは、圧倒的多数が「は」を含まない形であった。
- ・ 表には記載していないが、BCCWJ の「雑誌」「新聞」「教科書」における検索結果は、いずれの形も 0 であった。これは、この言い回しを学ぶ機会が稀である可能性が高いことをうかがわせる。この点は、法教育の観点から検討の必要があろう。

表 2 「この限りで(は)ない。」

	「少納言」(法律 SC)		MIC データ (法律)	
	件数	割合	件数	割合
この限りでない。	281	100.0%	1,506	99.7%
この限りではない。	0	0.0%	4	0.3%

### 5.2.2 「ものである(。)」

法律分野でよく見られる言い回しの例として、しばしば挙げられる表現である(判決におけるこの表現について言及するものとして、田中(2012))。検索結果を表 3 に示す。

表 3 「ものである(。)」

	「少納言」 (法律 SC)		MIC データ (法律)		MIC データ (政省令等)	
	件数	割合	件数	割合	件数	割合
ものであること。	168	46.4%	692	43.3%	2,486	56.5%
ものであること…	24	6.6%	145	9.1%	409	9.3%
ものであるとき	69	19.1%	367	23.0%	404	9.2%
ものである場合	55	15.2%	299	18.7%	580	13.2%
ものである。	0*	0.0%	5*	0.3%	0	0.0%
その他	46	12.7%	127	7.9%	525	11.9%
(合計)	362	—	1,598	—	4,404	—

(\*)脚注 13 で述べた 1 件を除外した値である。

法律 SC と MIC データ (法律) とでは、出現形の比率は大きくは異なることが分かる。一方、MIC データ (法律) と MIC データ (政省令等) とでは、出現形の比率の分布が少し異なる。法律と政省令とでは、規定される対象・内容が異なるが<sup>30</sup>、その点と関係がある可能性がある。この点は今後の検討課題の一つである。

また、MIC データ (法律) には「ものである。」が 5 件あるが、これはいずれも、法律の前文の文字列であった。<sup>31</sup>

<sup>30</sup> 法律は、国民の権利・義務に直接関わる事項を定めるのに対して、政省令は、法令から委任を受けた事項や法令の施行のために必要な技術的な細目を定めるというように、対象分野・内容がやや異なる。

<sup>31</sup> 感染症の予防及び感染症の患者に対する医療に関する法律、ハンセン病療養所入所者等に対する補償金の支給等に関する法律など。

## 6. おわりに

本稿では、「法文」の言語の分析のためにコーパスを用いる際の前提事項を整理しつつ、「法文」の代表格としての「法律」の分析の例を挙げた。ここで整理した内容を踏まえて、今後、「法文」の具体的な特徴の分析を進めていきたい。

## 文 献

John Gibbons (1976) *Forensic Linguistics : an Introduction to Language in the Justice System*

(邦訳『法言語学入門 司法制度におけることば』, 東京外国語大学出版会, 2013)

金子宏, 新堂幸司, 平井宜雄(2008)『法律学小辞典 第4版補訂版』有斐閣

末川博, 他(1991)『新法学辞典』日本評論社

鈴木達也(2003)「経過規定と旧法令の効力」立法と調査 237号

(参議院法制局「法制執務コラム集」として次のURLに転載されている。

<http://houseikyoku.sangiin.go.jp/column/column051.htm>)

田中伊式(2012)「『この事件は、～したものです』などの表現をめぐって」放送研究と調査 2012年5月号, pp.72

(<https://www.nhk.or.jp/bunken/summary/research/kotoba/037.html> よりダウンロード可能)

法制執務研究会(2007)『ワークブック法制執務』ぎょうせい

前川喜久雄(2013)「コーパスの存在意義」『講座日本語コーパス1 コーパス入門』, pp.1-31, 朝倉書店

町田顕(1980)「法文の長さ」判例タイムズ, 400号, pp.38

松田謙次郎(2011)「法令の言語変異を探る」トークス, 14号, pp.23-43

(<http://ci.nii.ac.jp/naid/110008095658> よりダウンロード可能)

松田謙次郎(2012)「法令に見られるサ変動詞の五段化・上一段化について 2001年から2011年のデータ分析」トークス, 15号, pp.37-48

(<http://ci.nii.ac.jp/naid/110008799628> よりダウンロード可能)

丸山ほか(2011)「『現代日本語書き言葉均衡コーパス』に含まれるサンプルおよび書誌情報の設計と実装」国立国語研究所内部報告書(LR-CCG-10-02)

矢野信(2013)「言語資料としての「判決文」の分析にまつわる問題点」, 本予稿集収録

## 関連 URL

法令データ提供システム (総務省) <http://law.e-gov.go.jp/cgi-bin/idxsearch.cgi>

日本法令索引 (国立国会図書館) <http://hourei.ndl.go.jp/SearchSys/>

衆議院規則 (衆議院) [http://www.shugiin.go.jp/index.nsf/html/index\\_houki3.htm](http://www.shugiin.go.jp/index.nsf/html/index_houki3.htm)

参議院規則 (参議院) <http://www.sangiin.go.jp/japanese/aramashi/houki/kisoku.html>

最高裁判所規則集 (最高裁判所) <http://www.courts.go.jp/kisokusyu/>

自治体 Web 例規集へのリンク集 (洋々亭) <http://www.hi-ho.ne.jp/tomita/reikidb/reikilink.htm>

官報情報検索サービス (独立行政法人国立印刷局) <https://search.npb.go.jp/>

少納言 (国立国語研究所) <http://chunagon.ninjal.ac.jp/>

中納言オンラインマニュアル <https://maro.ninjal.ac.jp/wiki/>