

令和6年度 0049-0226
「自然言語処理に適した法令表記に関する調査研究」
最終報告プレゼンテーション資料

2025年3月28日

本調査研究は、下記2つのテーマについて、実施しました。

(1) 自然言語処理手法に適した**法令中の併合的接続詞・選択的接続詞の表記の調査・実証**

法令中の併合的接続詞として用いられる「及び」・「並びに」、選択的接続詞として用いられる「又は」・「若しくは」については、どの語句とどの語句を論理積又は論理和として接続しているのかが分かりづらく、人間可読性及び機械可読性の観点から課題がある。こうした併合的接続詞・論理的接続詞で接続された語句の情報の階層性について、法令データベースにおいて人間可読性の高い方法で表現を行うのに資するため、

- ①**自然言語処理の手法により機械判読で分析を行うための手法を調査・実証し、**
- ②**自然言語処理を困難にする法令表記ルールの有無や有る場合の課題類型について整理分析**を行う。

(2) 自然言語処理手法に適した**法令中の定義語・略称表記の調査・実証**

法令中の条文において、定義語・略称を使用する際、「以下「○○」という。」等のように表記する。法令の立案段階においては、定義語・略称を含む法令を立案・改正する場合には、条文内の定義語・略称の表記の適正化を図る観点から、定義語・略称の定義を行った条文と、定義された定義語・略称を使用されている条文を、手作業による検索・目視確認など、業務負担が高い方法で把握・管理している点で課題がある。こうした定義語・略称に関して、自然言語処理の手法により、法令データベースにおいて、

- ①**定義語・略称を定義した条文から同一法令内で定義語・略称が使用されている条文へのリンク設定、**
- ②**定義語・略称が使用されている条文から定義語・略称を定義した条文へのリンク設定を可能にする手法について、調査・実証を行い、**
- ③**これらのリンク設定を困難にする法令表記ルールの有無や有る場合の課題類型について整理分析**を行う。

本調査研究では、本調査キックオフ時点の次の法令データを用いて実施しました。

e-Gov法令検索からダウンロード可能な全法令XMLデータ 11,508法令のうち

- 未施行を除外（11,508法令→9,229法令）
- 文語体で書かれている法令を除外（9,229法令→8,956法令）

8,956法令データを用いて実施

（内訳）

- 憲法 1件
- 勅令 14件
- 法律 2,020件
- 政令 2,359件
- 府省令 4,146件
- 規則 416件

日本語は、以下の特徴から、その文意や文脈を機械判読することが困難な言語です。

- 語の間にスペースを持たないため、語の境界を判別することが困難
- 語順が固定されておらず、語を入れ替えても意味が通じる
- 係る語と係られる語の間に無限に修飾語を入れることができる

このため法令条文を機械処理するためには、条文のテキストを形態素解析し、意味を持つ最小単位の語に分解するとともに、その語がどのような品詞であるのか、どの語に対して係り受け関係を持つのかという情報を付与する必要があります。

このような「語の分割」から「品詞情報の付与」「語の係り受け解析」の一連の処理を「自然言語処理」といいます。

本調査研究では、調査・分析を進める中で法令条文を自然言語処理し、品詞情報・係り受け情報を得る必要があったため、自然言語処理用オープンソースソフトウェア・ライブラリ spaCy、spaCyをベースとした日本語向け自然言語処理オープンソースソフトウェア・ライブラリ GiNZA 等以下のソフトウェアを使用して「表記パターン解析ツール」を開発し利用しました。

- anaconda: 24.11.3
- python: 3.12.7.final.0
- spaCy: 3.7.2
- GiNZA(ja-ginza-electra): 5.2.0
- streamlit: 1.40.1

開発したツールは、入力された文章に対し「分かち書き」「品詞情報の付与」「係り受け解析」の各処理がなされ、法令条文内でそれらの結果を利用した「併合的接続詞・選択的接続詞が用いられる表現」や「定義語・略称に関連する表現」のパターン抽出プログラムや、係り受け情報から文章内容を把握するためのプログラムのロジックに関する検討等に活用しました。

法令データの併合的接続詞・選択的接続詞の表記調査を実施するにあたり、日本語法令文をその品詞ごとに分類し、さらに係り受けを分析する必要があり、それらの表記パターンを解析するツールを開発し、実証実験の基本情報として用いることとしました。

localhost:8501

解析するテキストを入力してください

氏名又は名称及び住所並びに法人にあっては、その代表者の氏名

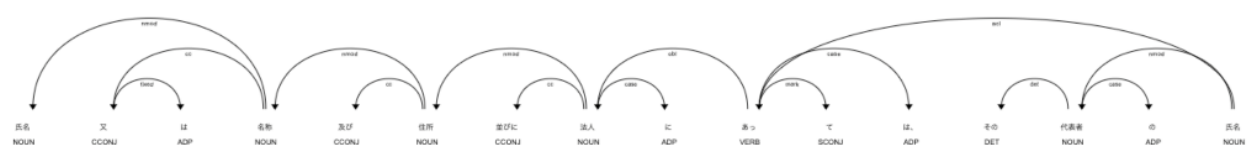
解析開始

解析結果

spacy + ginzaによる解析結果です

	text	lemma_	pos_	tag_	dep_	head_text
0	氏名	氏名	NOUN	名詞-普通名詞一般	nmod	名称
1	又	又	CCONJ	接続詞	cc	名称
2	は	は	ADP	助詞-係助詞	fixed	又
3	名称	名称	NOUN	名詞-普通名詞一般	nmod	住所
4	及び	及び	CCONJ	接続詞	cc	住所
5	住所	住所	NOUN	名詞-普通名詞一般	nmod	法人
6	並びに	並びに	CCONJ	接続詞	cc	法人
7	法人	法人	NOUN	名詞-普通名詞一般	obl	あつ
8	に	に	ADP	助詞-格助詞	case	法人
9	あつ	ある	VERB	動詞-非自立可能	acl	氏名

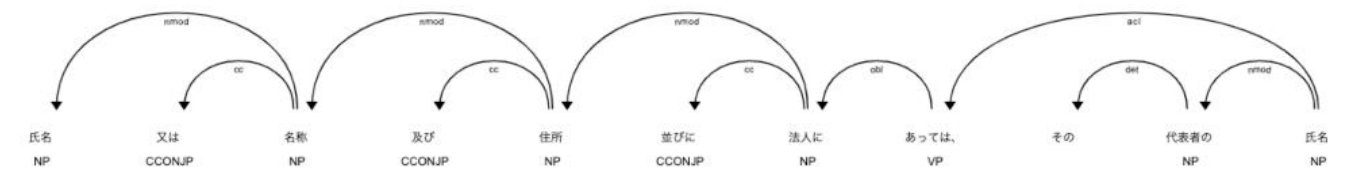
係り受け解析



The diagram shows a dependency parse tree for the sentence. The root node is 'acl' (attributive clause), which connects to '氏名' (NOUN) via 'acl'. The 'acl' node also connects to 'あつ' (VERB) via 'acl'. The 'あつ' node connects to '氏名' (NOUN) via 'acl'. The '氏名' node connects to '又' (CCONJ) via 'cc'. The '又' node connects to 'は' (ADP) via 'fixed'. The 'は' node connects to '名称' (NOUN) via 'nmod'. The '名称' node connects to '及び' (CCONJ) via 'cc'. The '及び' node connects to '住所' (NOUN) via 'nmod'. The '住所' node connects to '並びに' (CCONJ) via 'cc'. The '並びに' node connects to '法人' (NOUN) via 'nmod'. The '法人' node connects to 'に' (ADP) via 'case'. The 'に' node connects to 'あつ' (VERB) via 'obl'. The 'あつ' node connects to 'その' (DET) via 'det'. The 'その' node connects to '代表者' (NOUN) via 'nmod'. The '代表者' node connects to 'の' (ADP) via 'case'. The 'の' node connects to '氏名' (NOUN) via 'nmod'.

localhost:8501

Deploy



ここから独自検証

分割処理

Rule-Based MatchingでCCONJを頼りに分割した処理結果

```

{
  0: "氏名"
  1: "又は"
  2: "名称"
  3: "及び"
  4: "住所"
  5: "並びに"
  6: "法人にあつては、その代表者の氏名"
}

```

TODO

- 意味的な単位で分節として分けたあとに四則演算のような優先順位でパースするかたちは可能か？
- 読点を前処理して読点の機能を明確にしなければ正しく処理できない
- 文節で分割すると接続機能をもつ読点はケアできる？
- パーレンや定義・略称の表現をケアする必要がある
- 構造木からCCONJの係り受け状態から処理するものも考えてみる
- 複合的に接続詞が使われている場合の処理順序は？

TODO: 依存構造解析でのアプローチ

- headとcconjだけで意味が通じる文章がつかれる？ (CCONJのheadが？のときがある)

併合的接続詞「及び」「並びに」

- 「及び」は常に最小単位の接続となる
- 並列をさらに細分化して結び付ける必要がある場合には「並びに」が使用される
- 3つ以上の名詞が接続される場合には「、」（読点）を使う
(e.g. 「A、B、C及びD」)
- 接続されるものが動詞・形容詞・副詞の場合はその語が2つでも「、」（読点）を使う
(e.g. 「輸出し、及び輸入する」)

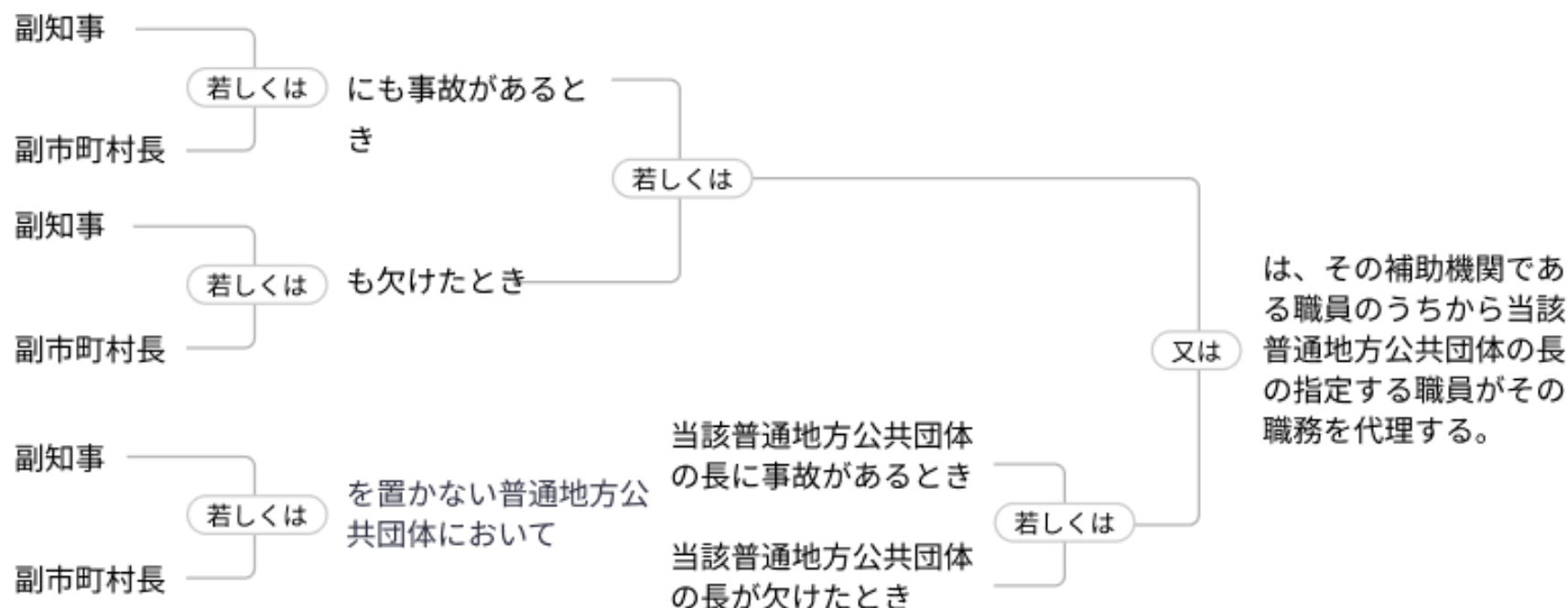
選択的接続詞「又は」「若しくは」

- 「又は」は常に一番大きな接続となる
- 選択をさらに細分化して結び付ける必要がある場合には「若しくは」が使用される
- 3つ以上の名詞が接続される場合には「、」（読点）を使う
(e.g. 「A、B、C又はD」)
- 接続されるものが動詞・形容詞・副詞の場合はその語が2つでも「、」（読点）を使う
(e.g. 「忌避し、又は同項の規定による」)

副知事若しくは副市町村長にも事故があるとき若しくは副知事若しくは副市町村長も欠けたとき又は副知事若しくは副市町村長を置かない普通地方公共団体において当該普通地方公共団体の長に事故があるとき若しくは当該普通地方公共団体の長が欠けたときは、その補助機関である職員のうちから当該普通地方公共団体の長の指定する職員がその職務を代理する。

?

副知事若しくは副市町村長にも事故があるとき若しくは副知事若しくは副市町村長も欠けたとき又は副知事若しくは副市町村長を置かない普通地方公共団体において当該普通地方公共団体の長に事故があるとき若しくは当該普通地方公共団体の長が欠けたときは、その補助機関である職員のうちから当該普通地方公共団体の長の指定する職員がその職務を代理する。



複雑な併合的接続詞・選択的接続詞を人が理解しやすくするために、自然言語処理により、条文を接続詞＋括弧＋読点等で区分し、表示するツールを開発し、検証しました。

JavaScriptのみによる接続詞分割プロトタイプ

職員の職務に係る倫理の保持を図るため、法律の規定に基づき内閣に置かれる各機関、内閣の統轄の下に行政事務をつかさどる機関として置かれる各機関及び内閣の所轄の下に置かれる機関並びに会計検査院並びに各行政執行法人（以下「行政機関等」という。）に、それぞれ倫理監督官一人を置く。

※処理後コンソールにArray(生データ)を表示しています

「接続詞＋丸括弧＋読点对応」モデル ▾

解析開始

クリア

処理結果

1行目

職員の職務に係る倫理の保持を図るため、法律の規定に基づき内閣に置かれる各機関、内閣の統轄の下に行政事務をつかさどる機関及び
内閣の所轄の下に置かれる機関並びに会計検査院並びに各行政執行法人（以下「行政機関等」という。）に、それぞれ倫理監督官一人を置く。

法令データに併合的接続詞・選択的接続詞はどれくらいあるのか

- 調査対象法令の全センテンス数： 3,439,036センテンス
- 併合的・選択的接続詞を含むセンテンス数：
975,859センテンス（全センテンスのおよそ28.4%）
- 併合的・選択的接続詞を含む「本則」のセンテンス数：
463,397センテンス（全センテンスのおよそ13.5%）
- 併合的接続詞「及び」「並びに」のみ使われている本則センテンス数：
184,965センテンス（全センテンスのおよそ5.4%）
- 選択的接続詞「又は」「若しくは」のみ使われている本則センテンス数：
202,652センテンス（全センテンスのおよそ5.9%）

複雑な併合的接続詞・選択的接続詞を条文化するのは、大変ではないか。

→接続詞を指定することで、成文化できないか。

→UIを検討し、検証ツールを開発した。

条文生成

先頭テキスト

代理人の死亡

代理人が

破産手続開始の決定

後見開始の審判

+

末尾テキスト

+

を受けたこと。

OR

OR

代理人の死亡又は代理人が破産手続開始の決定若しくは後見開始の審判を受けたこと。

第一編の輸入並びに第二編の手数料及び手数料の算付並びにこれらの徴収金の徴収又は算付に関する書類の送達及び公示送達については、地方税の例による。



先ほどの成文化ツールで見た課題

- 読点による接続判定
 - 名詞が3つ以上並ぶ場合
(地上権、永小作権、質権、使用貸借による権利、賃借権若しくはその他の使用及び収益を目的とする権利)
 - 動詞・形容詞・副詞が2つ以上並ぶ場合
(当該許可に係る掘削の工事を完了し、又は廃止したときは)
- 品詞解析の必要性
- 要素のパターンマッチングによる処理 が必要
- 複雑な親子構造を持つ条文を作成する際の支援
 - 入力に際し、あらかじめ各要素がそれぞれのレベルで接続されるのかを把握しておかなければ正しい入力が困難
- 文の要素を先に入力（登録）し、ドラックして構造を指定する等の工夫
(UI検討時に必要と認識していたが、プロト版に実装まで至らなかった)

(2) 法令中の定義語・略称表記の調査・実装

以下「〇〇」という。といった定義語・略称表記について類型分析を実施しました。

類型	内容	説明
パターンA	自然言語処理による抽出とXMLデータ構造を利用して処理できるもの	列挙パターン
パターンB	多少複雑な自然言語処理が必要となるもの	- かぎ括弧での文章パターン（ただし書きなどもある） - かぎ括弧なしの文章パターン（略称の定義も含まれる場合あり）
パターンC	複雑な自然言語処理や法制執務的な観点からのメタ情報が必要となるもの	- 定義条文等で定義されるが、その後、別の場所で再定義されるようなもの - 先に定義語が出現し、その後その定義がされるようなもの - 条件が先立つパターン - 他の法令での定義を使用するパターン
パターンD	表記方法や内容解釈を理由として対応するハードルが高いもの	- 除外条件があるもの（「～条は除く」のように「出現位置のような文脈的な除外」と捉えられるものと「～する者に限る」「～の場合は除く」のように「法解釈上での除外」と捉えられるもの） - 括弧書きの入れ子構造や並列配置などの表現（定義中の定義、定義と有効範囲を同じ階層で、定義と有効範囲を並列の括弧書き等）
パターンE	これまでの分析に加えてさらなる分析が必要であるもの	見出しなどで定義が含まれていることが示されていないもの

(2) 法令中の定義語・略称表記の調査・実装

類型分析は、e-Gov法令検索HTMLデータのリンク情報から機械抽出後、目視によるチェックを実施し、定義語・略称表記データを作成しました。

(2) 法令中の定義語・略称表記の調査・実装

それらの解析データを用いて、条文表記にそれらを表現するUIを検討し、テスト実装しました。



(2) 法令中の定義語・略称表記の調査・実装

二 前号に掲げる団体又は労働組合法（昭和二十四年法律第七十四号）第二条及び第五条第二項の規定に該当する労働組合が主体となつて構成され、自主的に労働条件の維持改善その他経済的地位の向上を図ることを主たる目的とする団体（団体に準ずる組織を含む。）であつて、次のいずれかに該当するもの

イ 一の都道府県の区域内において組織されているもの

ロ イ以外のものであつて厚生労働省職業安定局長（以下「職業安定局長」という。）が定める基準に該当するもの

（法第五条の三に関する事項）

第四条の二 法第五条の三第三項の厚生労働省令で定める場合は、次のとおりとする。

一 求人申込みをした公共職業安定所、特定地方公共団体若しくは職業紹介事業者の紹介による求職者、募集に応じて労働者になろうとする者又は供給される労働者（以下この項において「紹介求職者等」という。）に対して法第五条の三第一項の規定により明示された従事すべき業務の内容及び賃金、労働時間その他の労働条件（以下「従事すべき業務の内容等」という。）の範囲内で従事すべき業務の内容等を特定する場合

二 紹介求職者等に対して法第五条の三第一項の規定により明示された従事すべき業務の内容等を削除する場合

三 従事すべき業務の内容等を追加する場合

2 法第五条の三第三項の厚生労働省令で定める事項は、次のとおりとする。

一 前項第一号の場合において特定する従事すべき業務の内容等

二 前項第二号の場合において削除する従事すべき業務の内容等

三 前項第三号の場合において追加する従事すべき業務の内容等

3 法第五条の三第四項の厚生労働省令で定める事項は、次のとおりとする。ただし、第二号の三に掲げる事項にあつては期間の定めのある労働契約（当該労働契約の場合があるものに限り。以下この項において「有期労働契約」という。）に係る職業紹介、労働者の募集又は労働者供給の場合に限り、第八号に掲げる事項にあつては二条第二号に規定する派遣労働者をいう。以下同じ。）として雇用しようとする場合に限るものとする。

一 労働者が従事すべき業務の内容に関する事項（従事すべき業務の内容の変更の範囲を含む。）

二 労働契約の期間に関する事項

二の二 試みの使用期間に関する事項

二の三 有期労働契約を更新する場合の基準に関する事項（通算契約期間（労働契約法（平成十九年法律第二百二十八号）第十八条第一項に規定する通算契約期間をいう。）又は有期労働契約の更新回数に上限の定めがある場合には当該上限を含む。）

三 就業の場所に関する事項（就業の場所の変更の範囲を含む。）

四 始業及び終業の時刻、所定労働時間を超える労働の有無、休憩時間及び休日に関する事項

五 賃金（臨時に支払われる賃金、賞与及び労働基準法施行規則（昭和二十二年厚生省令第二十三号）第八条各号に掲げる賃金を除く。）の額に関する事項

六 健康保険法（大正十一年法律第七十号）による健康保険、厚生年金保険法（昭和二十九年法律第十五号）による厚生年金、労働者災害補償保険法（昭和二十二年法律第五十号）による労働者災害補償保険及び雇用保険法（昭和四十九年法律第十六号）による雇用保険の適用に関する事項

七 労働者を雇用しようとする者の氏名又は名称に関する事項

八 労働者を派遣労働者として雇用しようとする旨

九 就業の場所における受動喫煙を防止するための措置に関する事項

従事すべき業務の内容等

四条の二第1項第1号

四条の二第1項第2号

四条の二第1項第3号

四条の二第2項第1号

四条の二第2項第2号

四条の二第2項第3号

四条の二第6項

四条の二第7項

新する

違法第

(2) 法令中の定義語・略称表記の調査・実装

下記類型のうち、赤字部分が自然言語処理の課題として分析しました（それぞれの詳細は、最終報告書に例示）。

類型	内容	説明
パターンA	自然言語処理による抽出とXMLデータ構造を利用して処理できるもの	列挙パターン
パターンB	多少複雑な自然言語処理が必要となるもの	- かぎ括弧での文章パターン（ただし書きなどもある） - かぎ括弧なしの文章パターン（略称の定義も含まれる場合あり）
パターンC	複雑な自然言語処理や法制執務的な観点からのメタ情報が必要となるもの	- 定義条文等で定義されるが、その後、別の場所で再定義されるようなもの - 先に定義語が出現し、その後その定義がされるようなもの - 条件が先立つパターン - 他の法令での定義を使用するパターン
パターンD	表記方法や内容解釈を理由として対応するハードルが高いもの	- 除外条件があるもの（「～条は除く」のように「出現位置のような文脈的な除外」と捉えられるものと「～する者に限る」「～の場合は除く」のように「法解釈上での除外」と捉えられるもの） - 括弧書きの入れ子構造や並列配置などの表現（定義中の定義、定義と有効範囲を同じ階層で、定義と有効範囲を並列の括弧書き等）
パターンE	これまでの分析に加えてさらなる分析が必要であるもの	見出しなどで定義が含まれていることが示されていないもの

(1) 法令中の併合的接続詞・選択的接続詞の表記の調査・実証

ここまでの調査・実証から、法令条文における表記ルールはある程度パターン化できるものの、その性質上内容の解釈が要なものも多く、自然言語処理の導入のみで対応できるものは現時点では限定的であることが見えました。

既存条文を自然言語処理による解析のみで、「接続詞を含んだ複雑な条文」を構造的に解釈することは、自然言語処理における技術的な点と法令表記ルールの複雑さという点においてまだまだ困難なものでした。

しかしながら、読点の付与ルールなどの法制執務に関連する表記的なルールをある程度包括した仕様実装、それに加え形態素解析や品詞付与などサーバーサイド対応可能な自然言語処理で対応できる範囲の検討を踏まえることによって、今回検証した「条文生成を行うUI」を段階的に実装しながら、法制執務においてどのような情報提示の方法や機能が品質向上・効率化において有用かを検証していくことが、自然言語処理的なアプローチを取り入れた「併合的・選択的接続詞を含んだ複雑な条文」の効率的かつ高品質な作成・解釈へ発展できると考えられます。

(2) 法令中の定義語・略称表記の調査・実装

今回は文字列処理によるパターンマッチングと自然言語処理を含めた機械処理での対応、つまり「手作業によるデータ調整」を省いた形で処理できる範囲を調査したものの、現時点では一部のパターンを除き、様々なパターンを網羅したものを提示することは困難であることがわかりました。

この点において、まずは機械的な処理で逐次的な形で情報提示ができる仕組みづくりからではなく、ある程度の準備期間を経てデータを構築するという方向性を前提に、自然言語処理を活かした「語の定義箇所の抽出」「定義語・略語と思われる箇所の抽出」をサポートできるようなツールを活用した手作業によるデータ構築が現実的な対応と考えられます。

また、すでにXML等に情報が付与された既存データがあれば、それらを解析し、扱うデータへ変換し、法制執務に活用できる形にすることもひとつの手段ではある。

このときにも自然言語処理的なアプローチが有用であることは想像できます。

自然言語処理を活用し、法制執務の業務フローの効率化・品質向上を目指す場合、自然言語処理に関する技術的な限界（特に辞書やコーパスの充実というところによることが大きい）や表記ルールのより広範にわたる整理分析に依存するところが非常に大きいことが明確になった。

そのため現時点では「自然言語処理を活用したシステムを開発して業務効率化を図る」よりも「自然言語処理を活用したデータ構築を行い、そのデータを利用して業務効率化を図る」といったほうがより現実的な取り組みとなることが整理できたと考えます。

具体的には「法制執務における内容解釈を補助するツール」のように自然言語処理によるデータ処理を行い、法令に多くのアノテーション情報を付すことが必要と考えます。

そのような取り組みから構築されたデータをさらに自然言語処理やAIなどの強化学習に取り入れることで「法制執務における内容解釈を補助するツール」の性能をさらに向上させ、「ツールからデータへ」「データからツールへ」といった再帰的な開発アプローチで法制執務における自然言語処理の活用を進めていくことが考えられます。

そしてその先には「自然言語処理を活用したシステムを開発して業務効率化を図る」道筋が見えてくるだけではなく、「文章のオブジェクト表現」を実現し、「Rules As Code」のようなより発展的な法令条文の在り方へ繋げることができると考えます。

アイビー・システム株式会社