

法令文の構造的書き換え

宇野真人

島津明

北陸先端科学技術大学院大学 情報科学研究科

{u-masato, shimazu}@jaist.ac.jp

1 はじめに

法令を社会というシステムの仕様書と考え、法令の検査、検証を行い、安心な電子社会を実現するためには、法令文書の言語解析が必要になる [1].

田中ら [2] は、法令文の言語処理のためにその生成機構を分析し、法令文はその条文が規定したい権利や義務等の内容を表す部分 (効果部) と、それらが成立する条件を示す部分 (要件部) とで構成されると述べた (図 1) .

中村ら [3] は、田中らの要件・効果構造に基づいて法令文を解析し、論理式へ変換するシステムを開発した. また、機械学習を用いた要件効果・構造の解析 [4] や、複数文の法令文からその論理構造を抽出する研究も行われている [5].

法令文は、その要件・効果構造を解析しても、国民年金法などでは依然として文の長さや複雑さは残ってしまう. これが論理式化の精度を下げる原因となっている. また、法令は一般人が読んでも容易に理解できるような記述がされることが望まれるが、専門的な法令用語を除いても、可読性の高い記述がされているとは言い難い.

法令文の解析や読解を容易にするために、要件・効果構造において要件部と効果部がそれぞれ複数の要件と効果から構成されているとき、それらを明示的に示すことが考えられる. その際に、ゼロ代名詞や要件と

<S>政府は、</S><R>障害若しくは死亡又はこれらの直接の原因となった事故が第三者の行為によって生じた場合において、給付をしたときは、</R><E>その給付の価額の限度で、受給権者が第三者に対して有する損害賠償の請求権を取得する.</E>

図 2: 要件効果構造のラベル付き法令文

効果がうまく分けられない部分などに対応する必要がある.

本稿は、法令文の解析精度と可読性を向上させることを目指し、法令文を要件効果を構成する要素に分け、要素をその役割により 4 タイプに分類し、格フレーム辞書を用いてゼロ代名詞解析も行うことで、要件や効果が明示的になるように法令文を構造的に書き換える方法を提案する.

本稿は、以下 2 節で法令文の構造的書き換えについて述べ、3 節でそれに基づく提案手法を述べる. 4 節で国民年金法に対して提案手法を適用した実験結果を報告する.

2 法令文の構造的書き換え

2.1 構造的書き換え

本稿では、法令文の構造的書き換えとは、「要件効果構造に基づく法令文を、各要件と効果をそれぞれを明示した複数文に書き換えること」とする.

例えば、図 2 の要件効果構造に分かれた法令文では、要件部 (<R>〜</R> で囲まれた部分) を見ると 2 つの要件から構成されている.

1. 障害若しくは死亡又はこれらの直接の原因となった事故が第三者の行為によって生じた
2. 給付をした

ここで主題部 (<S>〜</S> で囲まれた部分) は、効果部の主語であるが、2 の要件の主語でもある. よって、上記の条文は図 3 のように書き換えられる. 要件・効果文は基本的に述語で終わるものとする².

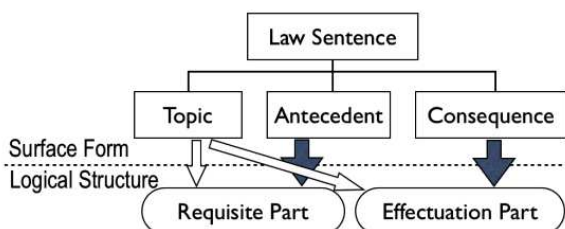


図 1: 法令文の標準構造と要件効果構造 [2] ¹

¹図 1 は [2] と同じではない. [2] では効果部がさらに対象部、内容部、規定部に分かれる.

²「前項の場合において、」のように他の条文を参照している表現は単独で参照文 [C] とした.

要件部:
R_1 : 障害若しくは死亡又はこれらの直接の原因となった事故が第三者の行為によつて生じた
R_2 : 政府は、給付をした
効果部:
E_1 : 政府は、その給付の価額の限度で、受給権者が第三者に対して有する損害賠償の請求権を取得する

図 3: 要件と効果を明示した法令文

市町村長 ϕ は、第一項又は第二項の規定による届出を受理したときは、厚生労働省令の定めるところにより、社会保険庁長官にこれを報告しなければならない。
↓
R_1 : 市町村長は、第一項又は第二項の規定による届出を受理した
E_1 : 市町村長は、厚生労働省令の定めるところにより、社会保険庁長官にこれを報告しなければならない

図 4: 法令文のゼロ代名詞補完

<S>[前項の規定によりその支給を停止するものとされた]E 年金給付の受給権者は、</S><E>同項の規定にかかわらず、その支給の停止の解除を申請することができる。</E>
↓
R_1 : 年金給付の受給権者は、前項の規定によりその支給を停止するものとされた
C : 同項の規定にかかわらず、
E_1 : 年金給付の受給権者は、その支給の停止の解除を申請することができる

図 5: 埋め込み文の要件化

2.2 ゼロ代名詞

格要素が省略されるゼロ代名詞は、法令文にも現れる。ゼロ代名詞が含まれる場合には、ゼロ代名詞を補完して要件・効果の項を明示する必要がある(図4)。法令文では、ゼロ化される名詞が限定し易いため、一般の文より高精度で解析できると考えられる。

2.3 埋め込み文

要件が連体修飾となり、その被修飾語が他の要件・効果の主語などになる場合がある。図5において、効果部の主語である主題部内の「前項の規定によりその支給を停止するものとされた」という埋め込み文は、要件とも考えられる。そのため連体修飾の要件がある場合には、図5のように、要件を抽出したい。また埋め込み文から要件を抽出する際には、被修飾語が要件の格要素であれば取る必要がある。

3 提案手法

3.1 構造的書き換えシステムの概要

要件効果構造のラベル付きの法令文を入力とし、図3のような複数の要件・効果文を出力する(図6)。処理は2つの部分に分かれる。1つは、法令文を要素に

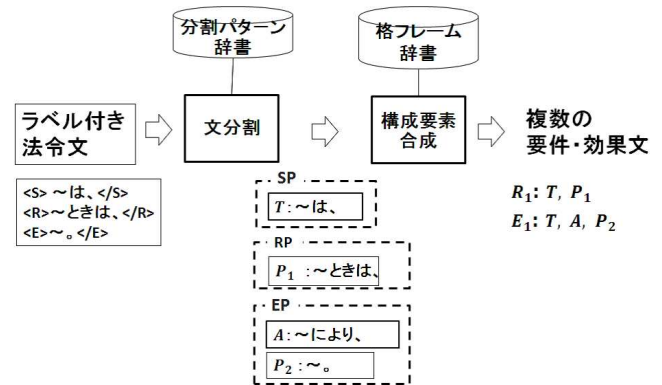


図 6: 書き換えシステムの処理の流れ

分割し、その役割によって分類を行う。前述した埋め込み文も構成要素に分割する。

要件・効果文は、構成要素の組み合わせで生成でき、また、その組み合わせ方は要素の主概念間の格関係で決められると考える。そのため次の処理で、各構成要素同士の組み合わせを判定し、合成によるゼロ代名詞の補完を行い、要件・効果文を出力する。

3.2 構成要素への分割

法令文では、読点の用法に明確な基準がある。永井ら[6]は、読点の用法と手がかり語に基づく法令文解析手法を示した。本研究でも、読点は法令文の要件効果を明確にするために重要な役割があると考え、句読点とその直前の表現によって分割できる部分を法令文を構成する要素とする。そして、分割するパターンによって、構成要素を4つのタイプに分類した(表1)。

タイプ主題[T]は、法令文の主題を表し、主に「～は、」で分割される。タイプ命題[P]は、要件・効果を示す述語を主とする要素である。タイプ項[A]は、タイプPの命題の項となる構成要素である。「について」などの格助相当句を含めて「格助詞+読点」で分けられる。分割パターン中の読点や格助詞などの直前の語を、その要素の主概念とする。

また「第一項の場合において、」のように、他の条文を参照している要素は、タイプ付加条件[C]とした。

表 1: 構成要素のタイプと分割パターン

構成要素タイプ	要素役割	分割パターン例	
主題 [T]	主題を示す	～は, ～とは, ～については	+ 読点
命題 [P]	要件・効果を示す	VP (連用形), VP 目に, VP には, VP ときは, VP 場合には, VP ものとする, VP することができる, ...	+ 句読点
項 [A]	命題の項となる	NP 格助詞 (格助詞相当句)	+ 読点
付加条件 [C]	他の条文を参照する	～の規定にかかわらず, ～の場合において, ～の他, ...	+ 読点

T: 六十歳未満の妻に支給する寡婦年金は、
→ P: 寡婦年金を、六十歳未満の妻に支給する
T: その寡婦年金は、

図 7: 埋め込み文の分割

3.3 埋め込み文の分割

2.3 節で見たように埋め込み文から要件を抽出する。本稿では、構成要素タイプ T と A の主概念が、埋め込み文により連体修飾されている場合のみを要件として抽出する対象とする³。対象の埋め込み文の被修飾語と述語である修飾語とで、法令用格フレーム辞書を用いて、述語の格スロットの深層格が agent, object, goal であった場合に、埋め込み文をタイプ P の要素としてその要素を分割する (図 7)⁴。

また、被修飾語が名詞句「A の B」や「A の B の C」の場合には、修飾語がどの名詞を格要素として取るかによって、分割点が変わる。そのため、「A の B」などがあるときは、述語の格スロットの用例と類似度を求め、その値が最大であった名詞の方を分割点とする (類似度の計算式は 3.4 節で)。

3.4 構成要素の合成

分割した構成要素を合成して要件・効果文を生成する。要件効果を作るタイプ P の要素 (E_P) と、タイプ T, A の要素 (E_T , E_A) との組み合わせを判定する。タイプ AC は、他の条文を参照しているため単独で参照文 [C] とした。 E_A と E_T の主概念は名詞、 E_P は述語である。 E_P と、 E_T , E_A の主概念間の格関係を判定し、格関係があればそれらの要素を合成する。

前から順に構成要素を見ていき、 E_P を見つけたら、その E_P の要素内で主概念である述語の格解析を行う⁵。 E_P の格フレームの空スロットに対して、それよ

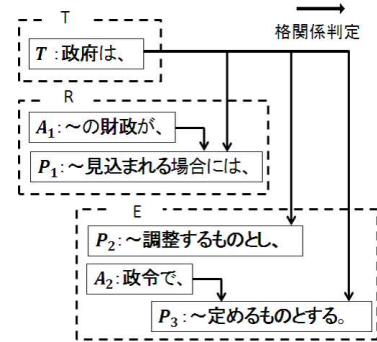


図 8: 構成要素の合成

り前にある要素との格関係を判定する。[6] によれば係り受けの制約により要件部の項が効果部へ係ることはない (主題は例外) ため、要件部の E_A と効果部の E_P との格判定は行わない (図 8)。また E_T と E_A の両方が E_P にかかる場合、 E_P に近い E_A から格判定を行い、格関係があればスロットに入れる。

格判定は、格フレームの表層格が一致しその用例との類似度を求めて閾値以上であれば格関係ありとした。類似度は以下の式で求める。

$$Sim(w_1, w_e) = \frac{2L}{l_1 + l_e} \quad (1)$$

ここで、 l_1, l_e はそれぞれ主概念 w_1 と用例 w_e のシソーラスの根からの深さ、 L は w_1, w_e の上位意味属性で一致している階層の深さを表す。閾値は 2 とした。

4 実験と評価

提案手法を国民年金法の各条項第 1 文に適用した。

4.1 埋め込み文の分割

要件が含まれる埋め込み文に対して、適切に要件が抽出できたかテストした。埋め込み文の分割を表 2 に示す。精度、再現率は、それぞれ 88%, 59% であった。要件化失敗と分析誤りの原因について以下に列挙する。

- 被修飾語が並列語であるとき、解析できなかった (2 個)
- 「～し、かつ、～する者」のように、修飾語が並列であるとき、前の要件の項が欠落した (13 個)

³例えば、「第一項の申出をした者に支給する老齢基礎年金の額は、」という部分から「(者は) 第一項の申出をした」という要件が取れるかもしれないが、埋め込み文が要件として取れるかどうかの選択は容易ではないので、本稿では対象外とする。

⁴法令用格フレーム辞書は研究室のものを用いる [3]。

⁵格解析は、形態素 [7]・構文解析結果 [8] に基づき行われる。詳細は [3] を参照。

表 2: 埋め込み文の要件化結果

	文数	%
要件化成功	46	57%
要件化されなかった	15	19%
項の欠落	13	16%
分割点誤り	6	8%
全埋め込み文数	80	100%
誤った要件化	3	

表 3: ゼロ代名詞補完結果

	ゼロ代名詞数	%
補完成功	55	82%
補完不成功	12	18%
全ゼロ代名詞数	67	100%
誤った補完	2	

- 判定詞「である」、「でない」は格フレームが存在せず、解析できなかった (5 個)
- 対応する格スロットの用例の誤り (7 個)

4.2 ゼロ代名詞の補完

ゼロ代名詞を含む文に対し、ゼロ代名詞が補完できるかテストした。実験結果を表 3 に示す。精度、再現率は、それぞれ 97%, 80% であった。ゼロ代名詞補完失敗の原因について、以下に列挙する。

- 対応する格フレームの不足 (6 個)
- 判定詞「である」は格フレームが存在せず、解析できなかった (2 個)
- 形態素・構文解析ミス⁶ (3 個)

4.3 議論

処理としては成功だが、可読性を高める要件として適切でないものは不正解にした。

その停止すべき期間の分として年金が支払われたときは、その支払われた年金は、その後に支払うべき年金の内払とみなすことができる。

↓

R_1 : その停止すべき期間の分として年金が支払われた
 R_2 : 年金は、その支払われた
 E_1 : その年金は、その後に支払うべき年金の内払とみなすことができる

⁶ 格助詞相当句を助詞として扱ったために、構文解析を誤ったものがある。

上の「年金は、その支払われた」という要件は、 R_1 の繰り返しだと考えられる。法令文では曖昧さを避けるため、文が長くなるときに語句などを繰り返す場合がある。この場合、冗長を避けるため、得られる要件が新しい要件か確認する処理が必要であろう。

国民年金法に現れる判定詞「である」のパターンを分析した。1. 「 N_1 が N_2 である (N_2 である N_1)」、2. 「 N_1 が VP (場合・とき・もの) である」、3. 「 N_1 が X (歳・日・月・円) + (以上・未満・後) である」の 3 パターンが見られた⁷。2 は「 N が VP する」として対応でき、1 は「被保険者である者」のように N_1 は N_2 と同性質だと考えられる。3 は「以上」などの前にある接尾辞(「歳」や「円」等)に着目すれば、 N_1 が限定できると考えられる。

5 おわりに

本稿では、法令文の要件効果構造を構造的に書き換え、要件・効果文として明示する手法について述べた。ゼロ代名詞の補完については、一般文に比べ高い精度で補完できた [9]。可読性の向上には、要件・効果を明示するだけでなく、四則演算を数式で表すなど文の表現を変えることが考えられる [10]。また、さらなる精度向上には並列構造の解析や、本稿では考慮しなかった但し書きや挿入文も考慮することが課題である。

謝辞 本研究の一部は、科学研究費補助金による

参考文献

- [1] 片山卓也. 検証進化可能電子社会 - 情報科学による安心な電子社会の実現. 情報処理, vol.46, pp. 515-512, 2005.
- [2] 田中規久雄, 川添一郎, 成田一. 法律条文の標準構造. 自然言語処理研究会, 97-12, 1993.
- [3] M. Nakamura, S. Nobuoka, and Akira Shimazu. Towards translation of legal sentences into logical forms. *JURISIN 2007*, pp. 6-17, 2007.
- [4] L. M. Nguyen, N. X. Bach, A. Shimazu. Supervised and semi-supervised sequence learning for recognition of requisite part and effectuation part in law sentences. *9th International Workshop on FSMNLP*, pp. 21-29, 2011.
- [5] K. Takano, M. Nakamura, Y. Oyama, A. Shimazu. Semantic analysis of paragraphs consisting of multiple sentences -towards development of a logical formulation system-. *JURIX*, pp. 117-126, 2010.
- [6] H. Nagai, T. Nakamura, Hirosato Nomura. Skeleton structure acquisition of japanese law sentences based on linguistic characteristics. *3rd NLPRS*, pp. 143-148, 1995.
- [7] S. Kurohashi, M. Nagao, T. Nakamura, and Y. Matsumoto. Improvements of japanese morphological analyzer juman. *International Workshop on SNLR*, pp. 22-28, 1994.
- [8] S. Kurohashi and M. Nagao. Kn parser : Japanese dependency/case structure analyzer. *International Workshop on SNLR*, pp. 48-55, 1994.
- [9] 飯田龍, 徳永健伸. 述語対の項情報を利用した文間ゼロ照応解析. 言処学会第 16 回年次大会, pp. 804-807, 2010.
- [10] 島津明. 国民年金法の構造的書き換え - 法令工学の立場から -. JAIST Press, 2009.

⁷ 1, 3 の N_1 は省略されることがある。